Perception multimodale de la distance dans un environnement virtuel

Nicolas Côté

Perception Multimodale de la Distance dans un Environnement Virtuel

Rapport d'activités

Nicolas Côté

- UFR Sciences & Techniques
- Laboratoire d'Informatique des Systèmes Complexes, E.A. 3883 (LISyC)
- 5 juillet 2011

TABLE DES MATIÈRES

ACRONYMES

ANOVA	Analyse de Variance *analysis of variance*
ARéVi	Atelier de Réalité Virtuelle
BRIR	Binaural Room Impulse Response
CAVE	Cave Automatic Virtual Environment
CERV	Centre Européen de Réalité Virtuelle
DRP	Drum Reference Point
ENIB	École Nationale d'Ingénieurs de Brest
ERP	Ear Reference Point
EV	Environnement Virtuel
HOA	Higher Order Ambisonics
HMD	Head-Mounted Display
HRTF	Head Related Transfer Function
ILD	Interaural Level Difference
ITD	Interaural Time Difference
KEMAR	Knowles Electronics Manikin for Acoustic Research
LCD	Liquid Crystal Display
LIMSI	Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur
LISyC	Laboratoire d'Informatique des Systèmes Complexes, E.A. 3883
OpenGL	Open Graphics Library
RIR	Room Impulse Response
UBO	Université de Bretagne Occidentale
VBAP	Vector Base Amplitude Panning
WFS	Wave Field Synthesis

INTRODUCTION

L'étude présentée dans ce rapport a été réalisée dans le cadre d'un programme du Conseil Général du Finistère pour l'accueil de post-doctorant ayant obtenu une thèse de doctorat à l'étranger et a été cofinancée par l'ENIB. Suite à ma soutenance de thèse qui a eu lieu le 30 Juin 2010 à l'université technique de Berlin, j'ai commencé ce post-doctorat d'une durée d'un an sous la responsabilité de Vincent Koehl et Mathieu Paquier. Cette section présente le programme de recherche ainsi que le laboratoire d'accueil.

Programme de recherche

La position des objets dans un espace est défini par trois coordonnées : l'azimut, l'élévation et la distance. Bien que certaines études aient étudié la perception d'objets sonores et visuels en élévation et en azimut, peu d'études se sont focalisées sur la perception en distance. Cependant, il a été montré que les humains sous-estiment la distance d'objets sonores et visuels distants et sur-estiment la distance d'objets sonores et visuels proches. L'humain perçoit la distance d'un objet et la profondeur d'une scène par le biais de plusieurs indices. Ceux-ci sont combinés afin d'estimer avec précision la distance et la profondeur. Le chapitre 1 introduit le processus de perception de la distance et définit les différents indices visuels et sonores de la distance.

Les applications de réalité virtuelle permettent de manipuler des objets dans un environnement généralement multimodal. Pour cela, il est nécessaire de reproduire fidèlement la position des objets dans cet environnement virtuel. La perception de la profondeur est affectée par différents facteurs inhérents aux environnements virtuels comme le champ de vision restreint ou la visualisation d'images de synthèse non réalistes. Le chapitre 2 décrit les différentes interfaces visuelles et sonores utilisées en réalité virtuelle et leur influence sur la perception de la distance et de la profondeur.

Une des techniques les plus employées en réalité virtuelle afin de reproduire la profondeur d'un environnement est la stéréoscopie. Cette technique reproduit certains indices de vision binoculaire. Cependant, cette technique ne permet pas de reproduire fidèlement l'ensemble des indices visuels car elle découple la convergence et l'accommodation. Cette étude a pour but de comparer, dans des tâches de perception de la distance égocentrique, la restitution sonore spatialisée et le rendu visuel stéréoscopique. Ainsi, l'étude décrite dans ce rapport analyse la perception de la distance dans trois contextes différents : modalité visuelle seule, modalité auditive seule, bimodalité. Le chapitre 3 décrit le protocole expérimental mis en place. Pour cela un environnement virtuel a été développé. Des stimuli sonores et visuels ont été disposés à des distances cibles de 2 à 20 m du sujet.

Résultats attendus

Lors de la définition du sujet de post-doctorat, l'étude se focalisait sur la définition d'une méthodologie adaptée à l'évaluation du son spatialisé lors de son utilisation dans un contexte de réalité virtuelle. Cependant, à la suite d'une étude bibliographique, cette étude s'est concentrée sur une dimension particulière du son spatialisé : la distance. En effet, peu d'études se sont focalisées sur la perception en distance or la vision 3D est de plus en plus utilisée dans les contenus audiovisuels. Cette technique permettant une meilleure restitution de la profondeur dans les scènes visuelles, il était intéressant de comparer les techniques de restitution d'indices de perception de la profondeur par le biais des deux modalités vision et audition. Cette étude a donné lieu à une publication dans une conférence internationale, voir Côté et al. [2011]. De plus, un article scientifique a été soumis dans la revue *Attention, Perception & Psychophysics*.

Actions de formation

Au cours de ce post-doctorat, j'ai apporté une expertise sur le son spatialisé à l'équipe de recherche spécialisée dans la réalité virtuelle. J'ai également présenté les résultats de cette étude au cours d'un séminaire organisé par le laboratoire d'accueil, le Laboratoire d'Informatique des Systèmes Complexes, E.A. 3883 (LISyC). Enfin, j'ai participé à quelques enseignements comme le programme *Méthodes et Pratiques Scientifiques* mis en place entre le Centre Européen de Réalité Virtuelle (CERV) et les lycées Kerichen et Vauban de Brest. Cette enseignement a permis d'initier les lycéens à la démarche scientifique en abordant notamment la réalité virtuelle et plus particulièrement la vision 3D et le son 3D.

Équipe de recherche

Le Laboratoire d'Informatique des Systèmes Complexes, E.A. 3883 (LISyC) regroupe des enseignant-chercheurs de l'Université de Bretagne Occidentale (UBO) et de l'École Nationale d'Ingénieurs de Brest (ENIB), effectuant leurs travaux de recherche dans le domaine de la réalité virtuelle, entre autres. En Janvier 2008, le laboratoire était composé de 45 permanents, 3 post-doctorants, 23 doctorants et 11 ingénieurs de recherche. Ceux-ci sont, pour certains, localisés au CERV, bâtiment regroupant différents chercheurs travaillant sur la thématique de la réalité virtuelle. Le groupe de recherche travaillant dans le domaine du son spatialisé est composé de :

- Mathieu Paquier, Maitre de Conférences, UBO,
- Vincent Koehl, Maitre de Conférences, UBO,
- Sylvain Marchand, Professeur, UBO (septembre 2011).

Les travaux actuellement menés par ce groupe de recherche portent sur l'évaluation de la qualité sonore des systèmes de restitution (enceintes acoustiques, casques) et d'enregistrement (microphone ambisonique).

Afin de réaliser cette étude, du matériel mis à disposition par le CERV a été utilisé :

- la salle immersive du CERV, voir section 3.1.2,
- une tête artificielle Neumann pour les enregistrements binauraux (calibration),

- un système de restitution sonore (projet MARVEST).

De plus, l'UBO a investi dans du matériel afin de réaliser l'expérience présentée dans le chapitre 3 :

- une carte son *Lexicon Alpha*,
- un casque *Sennheiser HD 650*.

Nicolas Côté :

Adresse :	Centre Européen de Réalité Virtuelle (CERV) 25, rue Claude Chappe BP 38 F-29280 PLOUZANÉ
Téléphone :	+33 (0)2 98 05 89 71
Fax :	+33 (0)2 98 05 89 79
E-mail :	nicolas.cote@univ-brest.fr

PERCEPTION MULTIMODALE DE LA DISTANCE

L'estimation précise de la distance des objets qui nous entourent est une tâche essentielle de notre vie. Par exemple, il est vital de connaître la position d'une voiture à l'approche lorsque nous nous apprêtons à traverser la rue. La perception du monde extérieur, au-delà de ce qui est à portée de la main, se fait essentiellement par deux sens : la vision et l'audition. Contrairement au toucher, ces deux sens donnent un aperçu de l'ensemble du monde et par conséquent ne donnent pas d'information précise sur l'environnement [Ménélas et al., 2009]. Plusieurs études ont montré que la vision permet une meilleure localisation des objets dans l'espace que l'audition [Loomis et al., 1998, Zahorik, 2001, Shelton and Searle, 1980]. Cependant, peu d'études se sont focalisées sur la perception en distance par comparaison aux nombres d'études portant sur la localisation en azimut ou en élévation.

Cette étude porte sur la perception en distance à partir d'indices visuels et auditifs. Après une introduction sur la perception en distance dans la section 1.1, ce chapitre présentera les processus de perception en distance pour la modalité auditive (section 1.2) et visuelle (section 1.3). Enfin, la section 1.4 présentera le processus d'intégration des différents indices visuels et auditifs pour obtenir une estimation précise de la distance.

1.1 Perception de la distance

Les humains sont capables de localiser les objets qui les entourent en leur attribuant une position dans l'espace. Pour définir la position de l'objet perçu (visuel ou auditif), plusieurs systèmes de coordonnées peuvent être utilisés : *cartésien*, *sphérique*. Nous pouvons définir un système de référence, basé sur le système de référence utilisé en anatomie, qui repose sur un ensemble de trois plans définis par rapport aux systèmes de perception, i.e. visuel ou auditif (voir figure 1.1).

- Le plan *horizontal* (transverse) : la source est au-dessus ou en-dessous.
- Le plan *médian* (sagittal, vertical) : la source est à gauche ou à droite.
- Le plan *frontal* (coronal) : la source est devant ou derrière.

Ces trois plans sont combinés à trois axes :

- Axe *rostro-caudal* : perpendiculaire au plan horizontal.
- Axe *droite-gauche* (interaural) : perpendiculaire au plan médian.
- Axe *dorso-ventral* : perpendiculaire au plan frontal.

Dans notre étude, nous utiliserons essentiellement le plan médian et l'axe interaural pour définir la position de l'objet perçu.

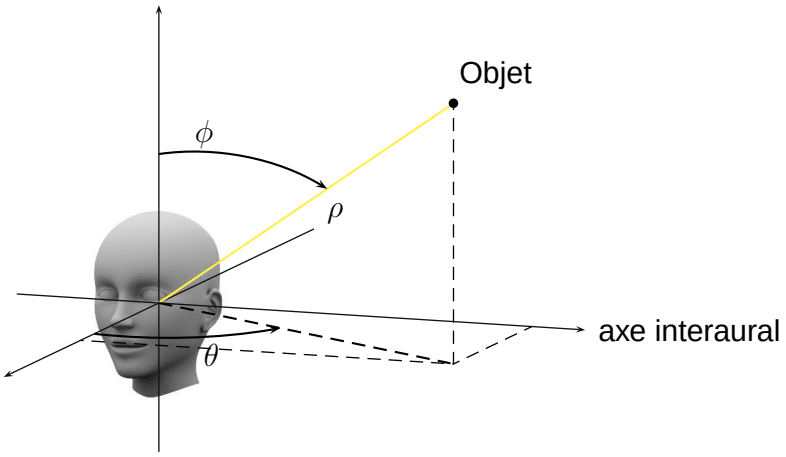


Figure 1.1 – Système de référence pour positionner un objet visuel ou auditif dans l'espace.

En se basant sur ces trois plans et trois axes, l'humain définit de manière plus naturelle la position d'un objet en utilisant un système de coordonnées sphérique plutôt que cartésien. Ces coordonnées sont l'azimut ($\phi \in 0 \dots 360^{\circ}$), l'élévation ($\theta \in 0 \dots 90^{\circ}$) et la distance ($\rho \in 0 \dots \inf$ m). Cette étude porte sur la perception en distance. On peut définir deux distances différentes :

- la distance *égocentrique* ou *absolue* séparant le sujet (le centre de sa tête) et l'objet perçu,
- la distance *relative* entre deux objets, on parle aussi de *relief*.

Plusieurs études ont montré que la distance égocentrique cible était sous-estimée, voir Da Silva [1985] et Philbeck and Loomis [1997] pour la vision et Zahorik et al. [2005] pour l'audition. Cet effet avait déjà été détecté par Békésy and Wever [1960]. La sous-estimation des grandes distances fut un processus pertinent pour nos ancêtres. En effet, la tendance à percevoir comme proche un danger, même si celui-ci est éloigné, nous permet de garder une "marge de sureté" vis-à-vis du danger.

Suivant Cutting and Vishton [1995], l'environnement visuel d'un observateur peut être divisé en trois sous-espaces : l'espace *personnel* (jusqu'à 2 m), l'espace d'*action* (entre 2 et 30 m) et l'espace *éloigné* (au-delà de 30 m), voir figure 1.2. La littérature montre que la perception en distance n'est pas identique pour chaque sous-espace. En effet, la distance égocentrique est :

- *sous-estimée* pour des objets situés dans l'espace d'action et l'espace éloigné,
- *sur-estimée* pour des objets situés dans l'espace personnel.

Ce phénomène apparait pour la modalité auditive [Zahorik, 2002a] et visuelle [Gogel and Tietz, 1973]. Ces deux espaces sont séparés par une distance spécifique. En effet, en absence d'indices de distance, le sujet positionne l'objet à une distance "par défaut", appelée *dark vergence* en perception visuelle, qui se situe à 1,9 m du sujet.

1.1.1 Méthodologie

La distance perçue est déterminée dans les expériences de perception par le biais de protocoles de mesure plus ou moins complexes. Une grande variété de protocoles expéri-

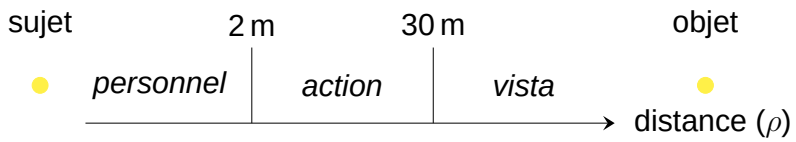


Figure 1.2 – Décomposition en trois sous-espaces de l'environnement visuel d'un observateur.

mentaux sont utilisés pour estimer la distance cible d'un objet dans l'espace. Par exemple, il peut être demandé aux sujets d'estimer la distance d'un objet :

- verbalement en utilisant une échelle interne, e.g. 12 m [Gogel and Tietz, 1973],
- à partir d'une estimation de grandeur sur une échelle de mesure, e.g. en utilisant un curseur [Teghtsoonian and Teghtsoonian, 1978],
- à partir d'une mesure relative entre deux objets, e.g. *A* 2 fois plus loin que *B* [Da Silva, 1985],
- via une action dirigée, e.g. par triangularisation [Loomis et al., 1998, Mershon and Hutson, 1991, Klein et al., 2009].

Cette dernière méthode utilise les mouvements du sujet humain pour localiser les objets dans l'espace. Par exemple la méthode par triangularisation se déroule en trois étapes :

1. le sujet, faisant face à l'objet à percevoir, le visualise ou l'entend,
2. il tourne puis marche, les yeux fermés, jusqu'à un point défini par l'expérimentateur ,
3. il localise la position de l'objet (par exemple en se tournant vers l'objet).

Même si toutes ces méthodes de mesure permettent de quantifier la distance égocentrique d'un objet, Da Silva [1985] montre qu'elles ont une influence sur la distance perçue. D'autres facteurs, comme l'environnement de test (e.g. intérieur ou extérieur), ou le choix des distances évaluées (champ proche ou champ lointain), ont également un impact sur les distances perçues. En effet, des méthodes d'évaluation indirectes comme l'utilisation d'actions dirigées permettent de réduire la sous-estimation des distances égocentriques cibles inférieures à 20 m [Andre and Rogers, 2006, Klein et al., 2009]. Il faut cependant noter qu'une méthode par triangularisation induit un mouvement du sujet et donc une variation des indices acoustiques et/ou visuels ce qui apporte un supplément d'informations [Ashmead et al., 1995, Speigle and Loomis, 1993].

1.1.2 Modélisation

Teghtsoonian and Teghtsoonian [1978] ont montré que la distance perçue pouvait être estimée de manière instrumentale en utilisant la loi de Stevens [1957] :

$$\rho' = k\,\rho^{a} \tag{1.1}$$

où ρ' représente la distance perçue, ρ est la distance réelle de l'objet (i.e. distance cible) et k et a sont deux coefficients. L'exposant a reflète la compression où l'expansion de la distance cible. Lorsque a est proche de 1, il existe une relation linéaire entre la distance perçue et la distance cible. Le coefficient k représente le facteur d'échelle entre l'espace perçu et l'espace physique.

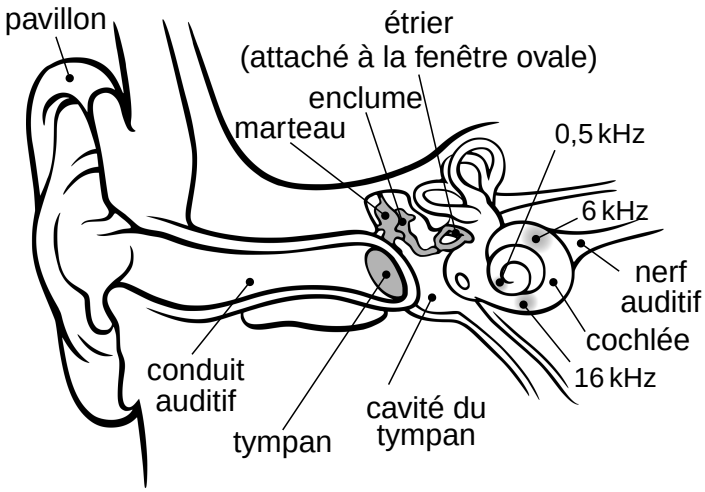


Figure 1.3 – Système auditif humain.

1.2 Perception auditive

Les humains perçoivent les sons de manière passive. Même si certains processus cognitifs permettent de focaliser son attention sur une source sonore particulière, l'ensemble des sons qui nous entourent sont perçus, quelle que soit leur position. L'onde acoustique qui se propage dans l'air arrive aux deux tympans après avoir subi quelques modifications dues à la forme du corps, de la tête et de l'oreille externe du sujet et surtout à la distance entre les deux oreilles. Ce sont ces modifications qui produisent les indices de localisation utilisés par l'auditeur pour estimer la position d'une source sonore. Mais, ces indices acoustiques ont également d'autres utilités comme :

- *Séparer* de multiples sources sonores simultanées (effet "cocktail-party").
- *Définir* les caractéristiques de l'environnement (taille, matériaux, ...) car celui-ci influence les indices acoustiques par le biais de réflexions du champ sonore direct sur les murs et les objets.

1.2.1 Le système auditif humain

Le système auditif humain transforme une onde acoustique en un influx nerveux, i.e. une série d'impulsions électriques, jusqu'au système nerveux central. La figure 1.3 présente le système auditif composé de l'oreille externe (pavillon, conduit auditif), moyenne (marteau, enclume, étrier) et interne (cochlée). Généralement, on considère que les fréquences entre 20 Hz et 20 000 Hz sont perçues. Cependant, cette moyenne varie suivant plusieurs critères, comme la fatigue ou l'âge de l'auditeur.

1.2.2 Les indices acoustiques de perception de la distance

Mershon and King [1975] ont défini quatre indices acoustiques de perception de la distance : l'intensité, la réverbération, le contenu spectral et les différences binaurales. Les deux premiers indices sont ceux qui varient le plus avec la distance et les plus importants d'un point de vue perceptif [Shinn-Cunningham, 2000a]. Cette section présente ces différents indices et leur influence sur l'estimation de la distance d'une source auditive.

Intensité

L'intensité est souvent considérée comme l'indice principal de perception de la distance d'une source sonore. En effet, la variation de cet indice en fonction de la distance est très bien détectée par un auditeur. Pour une onde acoustique traversant un champ libre, son énergie est inversement proportionnelle à la racine carrée de la distance, ρ, entre la source sonore et le récepteur (i.e. proportionnel à $1/\rho^2$). Pour une source sonore relativement éloignée ($\rho > 1$ m) et en champ libre, une doublement de la distance de la source introduit une réduction de 6 dB de l'énergie reçue [Coleman, 1963]. Cependant cette loi n'est pas applicable pour une source sonore située en champ proche, i.e. $\rho < 1$ m. En effet, dans ce cas, la présence de la notre tête influence le niveau sonore au niveau des deux oreilles.

Cependant, sans connaître à priori le niveau de la source sonore, l'intensité n'apporte qu'une information restreinte de la distance de la source sonore. En effet, le niveau sonore perçu par l'auditeur est relatif au niveau d'émission de la source sonore. Ainsi, il est difficile d'estimer la distance égocentrique d'une source sonore en champ libre [Gardner, 1969, Mershon and King, 1975]. De plus, la relation entre la distance cible et la distance perçue à partir du seul indice intensité, semble montrer une compression de la distance cible (i.e. $a < 1$) [Cochran et al., 1968, Begault, 1991].

Enfin, l'intensité est un paramètre physique et non perceptif. Ainsi, Stevens and Guirao [1962] ont montré que la distance perçue avait une relation linéaire avec la sonie, i.e. $a = 1$, qui est un paramètre perceptif lié à l'intensité.

Réverbération

La perception de la distance entre la source sonore et l'auditeur dépend en grande partie du ratio entre l'énergie du champ direct, I_d, et celle du champ réverbéré, I_r, i.e. les réflexions introduites par la salle d'écoute. Ces réflexions sont déterminées par la forme et les propriétés acoustiques des murs de la salle d'écoute et également par les objets situés dans la salle d'écoute. Ainsi le rapport champ direct/champ réverbéré (I_d/I_r) peut donc être vu comme un indice absolu de la distance. En effet, pour une source sonore lointaine ($\rho > 1$ m), le niveau du champ réverbéré est supposé indépendant de sa position dans la salle d'écoute. Par conséquent, le rapport I_d/I_r est inversement proportionnel au carré de la distance. La réverbération peut être définie par le temps de réverbération, T_{60}, défini en ms, qui correspond au temps de décroissance de l'intensité sonore de 60 dB.

Mershon and King [1975] ont montré qu'un auditeur estime plus précisément la distance égocentrique cible dans un environnement réverbérant qu'anéchoïque. Cependant, une augmentation de la réverbération du lieu d'écoute augmente la distance perçue de la source sonore [Butler et al., 1980, Nielsen, 1993, Mershon and Bowers, 1979]. En se basant sur ce phénomène, Bronkhorst and Houtgast [1999] ont développé un modèle permettant d'estimer la distance perçue à partir du rapport entre le champ direct et le champ réverbéré (coefficient de corrélation de 0,94 pour la base de données d'entraînement). Cependant, cet indice semble moins précis que l'intensité pour estimer la distance relative, ou pour détecter que la position d'une source sonore ait changée.

Enfin, Shinn-Cunningham [2000b] a montré que les auditeurs étaient capables d'adapter leur perception en fonction de la salle d'écoute. En effet, c'est l'expérience et l'apprentissage de la réverbération qui permet aux sujets de positionner avec précision un objet sonore dans l'espace [Kopčo et al., 2004].

Contenu spectral

D'après Blauert [1997], pour un objet situé à une distance supérieure à 15 m, les propriétés de propagation dans l'air de l'onde acoustique modifient significativement le contenu spectral du son. Cette absorption de l'air peut être modélisée par l'équation suivante [Coleman, 1963] :

$$\alpha_{air,f} = \exp(-2{,}25 \cdot 10^{-4} \frac{50}{h} \cdot f^{1,7}) \tag{1.2}$$

où h correspond au pourcentage d'humidité dans l'air et f correspond à une fréquence en Hz. Cette équation montre que l'absorption de l'onde acoustique dans l'air atténue principalement les hautes fréquences [Coleman, 1968]. A l'inverse, un signal avec beaucoup de hautes fréquences sera perçu comme proche de l'auditeur. Il est important de noter que la réverbération peut également avoir un impact sur le contenu spectral. En effet, les propriétés acoustiques des murs de la salle d'écoute ont une influence sur la coloration du signal sonore (i.e. balance entre fréquences aiguës et fréquences graves).

Comme pour l'intensité, sans connaître à priori le spectre du son, la coloration introduite par la propagation du son dans l'air n'apporte qu'une information restreinte sur la position de la source sonore.

Indices binauraux

Lorsque la source sonore n'est pas positionnée devant l'auditeur, i.e. $\theta = 0^{\circ}$et $\phi = 0^{\circ}$, les indices acoustiques arrivant aux deux oreilles de l'auditeur définissent l'azimut et l'élévation de cette source sonore. La théorie appelée *duplex*, proposée par Rayleigh [1907] et étendue par Blauert [1997], définit un principe de base de la localisation en azimut et en élévation d'une source sonore. Cette théorie modélise la tête humaine par une simple sphère de diamètre, r, compris entre 6 et 10 cm. Un son provenant de côté, avec un azimut θ, induit des différences acoustiques entre le signal arrivant à l'oreille *ipsilatérale* (la plus proche de la source sonore) et celui arrivant à l'oreille *controlatérale* (le plus éloignée de la source sonore) dues à la distance entre les deux oreilles, i.e. $d = 2r \cdot \sin(\theta)$, voir figure 1.4. Deux indices définissant l'azimut de la source sonore peuvent être déduits de ces différences acoustiques :

- Une *différence de phase*, Δt, ou Interaural Time Difference (ITD) :

  Le son va d'abord atteindre l'oreille *ipsilatérale* avant d'atteindre l'oreille *controlatérale*. Cependant, cet indice est perçu lorsque la longueur d'onde, λ, est supérieure à la distance entre les deux oreilles. Pour $\lambda < 0.2$ m, soit $f > 1500$ Hz, la distance entre les deux oreilles dépasse la longueur d'onde. Dans ce cas, l'ITD n'apporte plus d'information à l'auditeur sur la position de la source sonore.

- Une *différence de niveau*, ou Interaural Level Difference (ILD) :

  Pour des fréquences supérieures à 1500 Hz, , la tête de l'auditeur est plus grande que la longueur d'onde et ainsi crée une *ombre acoustique* (i.e. introduit des réflexions,

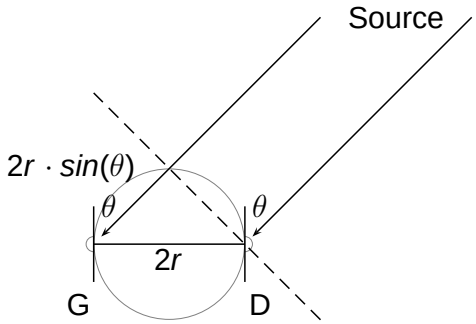


Figure 1.4 – Modélisation de la tête de l'auditeur suivant la théorie *duplex*, proposé par Rayleigh [1907].

résonances, et diffractions) qui atténue l'onde acoustique au niveau de l'oreille *controlatérale*. Ainsi, le son est plus fort au niveau de l'oreille *ipsilatérale* qu'au niveau de l'oreille *contralatérale*. Cependant, pour $\lambda > 0.2$ m, soit $f < 1500$ Hz, l'ILD est quasiment imperceptible.

L'onde acoustique arrivant aux tympans est également influencée par les caractéristiques physiques de chaque individu comme la forme de la tête et de l'oreille externe (pavillon). Ces caractéristiques introduisent un filtrage des indices spectraux, qui permettent de déterminer l'élévation de la source sonore. L'ILD, l'ITD et les indices spectraux sont caractérisés par une fonction appelée Head Related Transfer Function (HRTF). Ces filtres HRTF sont définis pour chaque oreille (H_G et H_D) et chaque position de la source sonore (θ, ϕ et ρ). Cependant, pour des distances supérieures à 1 m, les filtres HRTF ne sont que peu influencés par la distance, ρ [Duda and Martens, 1998].

Lorsque la source sonore est proche de l'auditeur et n'est pas juste devant l'auditeur (i.e. $\theta = 0^\circ$et $\phi = 0^\circ$), l'auditeur est capable d'extraire une information de distance à partir de ces deux indices ITD et ILD [Coleman, 1968]. Contrairement au niveau sonore global les différences binaurales permettent une évaluation de la distance de la source de manière absolue. Ces différences binaurales sont particulièrement efficaces dans un environnement anéchoïque mais n'apportent aucune information supplémentaire pour des conditions réverbérées ou pour des distances égocentriques supérieures à 1 m. En effet, pour des distances inférieures à 1 m, ce sont les réflexions, résonances et diffractions au niveau de la tête de l'auditeur qui introduisent ces effets binauraux.

Indices dynamiques

Les indices de perception en distance décrits dans les sections précédentes varient en fonction de la position de la source sonore. Ces variations dues à la translation ou la rotation de cet source sonore augmentent la quantité d'information disponible. Cependant, ces mouvements induisent trois nouveaux indices :

- le *parallaxe de mouvement* qui induit un changement de direction de la source sonore et une modification des indices acoustiques,
- le τ *acoustique* qui définit la durée que va mettre l'objet pour toucher l'auditeur,

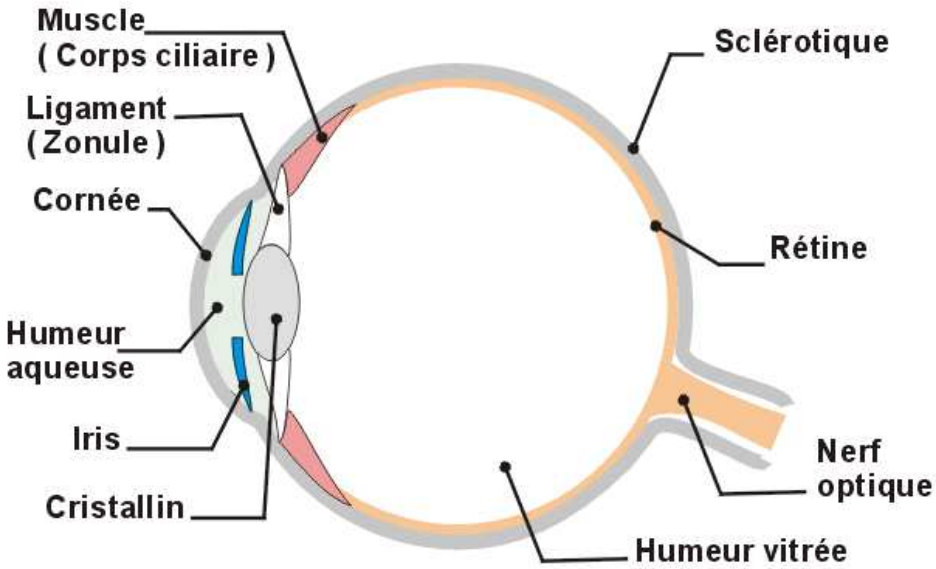


Figure 1.5 – Système visuel humain.

- l'effet *Doppler*, qui introduit une variation continue du spectre de la source sonore lors des mouvements de la source sonore. Les fréquences augmentent lorsque la source sonore se rapproche de l'auditeur et elles diminuent lorsque la source sonore s'éloigne de l'auditeur. Cependant cet effet n'a qu'une faible influence sur la perception de la distance [Rosenblum et al., 1987].

Enfin, la localisation en azimut et élévation d'une source sonore est influencée par les mouvements de l'auditeur [Majdak et al., 2008]. Ces mouvements sont perçus par les sens vestibulaire (lié à l'équilibre) et proprioceptif (lié aux mouvements du corps et de ses parties). Il est intéressant de noter que le τ acoustique est également sous-estimé par les auditeurs [Ashmead et al., 1995] ce qui est cohérent avec la sous-estimation de la distance égocentrique.

1.3 Perception visuelle

La vision humaine est, au contraire de l'audition, perçue activement. En effet, les yeux sont presque toujours en mouvement : ils convergent vers un point de fixation grâce aux muscles extra-oculaires et accommodent sur ce point grâce aux cristallins. La section 1.3.1 introduit les bases anatomiques et physiologiques du système visuel humain. Ensuite, les indices monoculaires et binoculaires de perception du relief et de la profondeur sont définis dans le section 1.3.2.

1.3.1 Le système visuel humain

Le système sensoriel de la vision transforme une onde électromagnétique en un influx nerveux, i.e. une série d'impulsions électriques, jusqu'au cortex visuel. Le système visuel humain est constitué de deux récepteurs, les yeux, présentés dans la figure 1.5. Les yeux sont placés symétriquement de part et d'autre du plan médian. Les images pour chaque œil, se projettent sur la rétine constituée des photorécepteurs : les cônes et les bâtonnets.

La distance entre le centre de la pupille des deux yeux est appelée distance inter-pupillaire. Cette distance est généralement comprise entre 50 et 70 mm pour la majorité des adultes, avec une moyenne de 63 mm. De plus, les yeux sont constamment en mouvement grâce aux muscles extra-oculaires. Pour percevoir une seule image, les deux yeux font converger leur axe visuel vers un point unique de fixation. En effet, la convergence des axes visuels en un point est nécessaire pour la fusion binoculaire. L'axe visuel passe par la pupille qui est le trou situé au milieu de l'iris.

Le champ visuel, ou *Field of View* (FoV), azimutal et vertical se défini en degré (°). Pour chacun des deux yeux, le champ visuel azimutal est d'environ 150° et se recouvre sur une large zone (environ 120°) dans laquelle se situe le point de fixation sur lequel convergent les axes visuels. Cette zone de 120° constitue le champ visuel binoculaire où la perception de la profondeur est plus précise. Ainsi, les deux yeux couvrent un champ visuel de 200°. Le champ visuel vertical est d'environ 135°.

1.3.2 Les indices visuels de perception de la distance

Le système visuel humain extrait les informations de profondeur et de position d'objet en distance à partir de nombreux indices. Ces indices sont perçus, traités par certains processus cognitifs, puis interprétés pour déterminer la position des objets dans la scène visuelle. Ces indices sont classés en deux grandes catégories :

- Les indices proprioceptifs (accommodation et convergence) sont ajustés par le système visuel. Ces indices sont fiables à faible distance (quelques mètres).
- Les indices visuels, composés des indices binoculaires et des indices monoculaires.

Le tableau 1.1 propose une liste exhaustive d'indices visuels définis par Cutting and Vishton [1995].

Tableau 1.1 – Liste des indices visuels de perception de la distance et de la profondeur suivant Cutting and Vishton [1995].

Catégorie		Indice
monoculaire	picturaux	perspective aérienne (brume), ombre, perspective linéaire, hauteur dans le champ visuel, taille relative (ou familière) gradient de texture, occultation
	cinétiques	parallaxe de mouvement, perspective de mouvement,
	accommodatif	accommodation
binoculaire		convergence, disparité

Indices proprioceptifs

Les indices proprioceptifs, convergence et accommodation, sont les indices liés à la perception active de la scène visuelle. Ceux-ci apportent une information précise de la profondeur principalement pour des objets situés dans l'espace personnel ($\rho < 2$ m).

- La *convergence* correspond à l'orientation de l'axe visuel (faisceaux lumineux partant de l'objet et arrivant à la rétine) de chaque œil vers une même cible, le point de fixation. Ainsi chaque œil perçoit une image légèrement différente de la même

scène visuelle. Ces deux images légèrement différentes d'un même objet permettent d'extraire une information de profondeur.

- L'*accommodation* est liée à la convergence. En effet, le changement de la convergence entraîne de façon réflexe une accommodation sur le nouveau point de fixation. l'accommodation est réalisée par la modification du cristallin. En absence d'objet visuel, un observateur a tendance à accommoder à une distance de 75 cm, appelée *dark focus* [Patterson, 2007].

Indices monoculaires

Il est important de noter qu'avec seulement un œil, l'homme interprète l'image reçue et en déduit des notions de profondeur par le biais d'indices monoculaires, voir figure 1.6. Ceux-ci sont appris inconsciemment, depuis le plus jeune âge. Cette section introduit les différents indices visuels monoculaires.

- L'*occultation* d'un objet par un autre permet de positionner relativement les objets en profondeur. L'occultation est un indice visuel de distance relative.
- La *taille* relative de l'objet est, comme pour l'intensité en audition, un indice principal de perception de la distance. En effet, la taille décroit avec la distance. Cependant, sans information sur les dimensions de l'objet, la taille reste un indice de perception de la distance relative.
- La finesse des textures donne une information complémentaire de la profondeur. La texture d'une surface est perçue plus nettement si la surface est positionnée à faible profondeur. C'est le *gradient de texture*.
- Les variations de lumière et les *ombres* sur les objets augmentent le "relief".
- En extérieur, la variation de la visibilité due au degré de transparence de la couche atmosphérique, appelé *épaisseur optique*, donne une information de profondeur sur les grandes distances (i.e. au-delà de 100 m). Elle est définie par la fraction de lumière diffusée ou absorbée par les composants de la couche traversée. Ces composants correspondent à des particules en suspension dans l'air (poussière et pollution).
- La *perspective* donne une information de l'espace tridimensionnel : e.g. sol, murs, plafond, objets. Cet indice, employé en peinture depuis l'époque de la Renaissance, permet de percevoir la profondeur dans une scène visuelle représentée en deux dimensions.

Vision binoculaire

La vision binoculaire correspond à l'extraction des informations de profondeur et de distance entre les objets et les surfaces d'une scène visuelle. Cette propriété, appelée *stéréopsis*, se base sur les différences entre l'image rétinienne de l'œil gauche et celle de l'œil droit. Ces différences sont induites par le couple convergence/accommodation (indices proprioceptifs) et la disparité binoculaire.

- La *disparité binoculaire* permet de définir la position relative des objets devant et derrière le point de fixation. Cette disparité correspond aux différences de position des projections d'un même objet sur les deux rétines. Lorsque l'objet se situe sur l'*horoptère*, un arc géométrique défini par l'ensemble des points de l'espace se projetant en des positions rétiniennes correspondantes, les deux images fusionnent pour

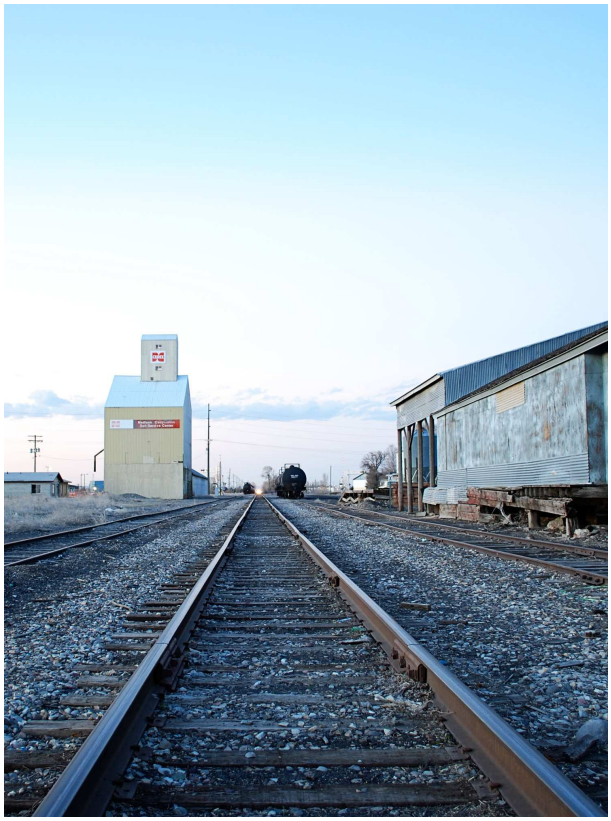


Figure 1.6 – Exemple d'image comprenant plusieurs indices monoculaires (perspective, gradient de texture) permettant une perception de la profondeur.

former une image de l'objet. Au delà d'une limite appelée aire de fusion de *Panum*, les deux images ne peuvent être fusionnées car la disparité est trop importante : c'est le phénomène de *diplopie*. Comme pour l'accommodation et la convergence, l'influence de la disparité binoculaire sur la perception de la distance s'atténue linéairement avec l'augmentation de la distance par rapport à l'observateur. Cependant, c'est l'indice le plus pertinent dans l'espace personnel (jusqu'à 2 m de distance). Pour une description plus détaillée de la perception visuelle de la profondeur, voir Patterson [2007].

De nombreuses études ont conclu que les indices binoculaires et proprioceptifs amélioraient la localisation en distance d'objets uniquement en champ proche, i.e. pour ρ inférieur à quelques mètres [Cutting and Vishton, 1995, Philbeck and Loomis, 1997]. Par exemple, Creem-Regehr et al. [2005] montrent que le type de vision (monoculaire ou binoculaire) d'une scène visuelle réelle n'a pas d'influence sur la performance de la localisation en distance (pour des distances cibles de 2 à 12 m). Gregory [1997] considère que les humains ont une vision monoculaire pour des distances supérieures à 6 m. En effet, l'accommodation et la convergence sont deux indices de perception de la distance efficaces en champ proche. Au-delà d'une certaine distance l'angle de convergence est quasi-nul. Cependant, certains auteurs ont montré que la disparité binoculaire est un indice de perception de la distance pertinent pour de grandes distances (20 et 40 m) [Palmisano et al., 2010].

Indices dynamiques

Comme pour l'audition, les indices visuels décrits dans les sections précédentes varient en fonction de la position de l'observateur. Ces variations dues à la translation ou la rotation de cet observateur augmentent la quantité d'informations disponibles. Les mouvements de l'observateur, ou des objets dans la scène visuelle, induisent deux nouveaux indices :

- Le *parallaxe de mouvement* correspond aux mouvements de l'observateur par rapport aux objets, ou les mouvements des objets entre eux. Ces mouvements peuvent être une translation ou une rotation. Cet indice induit un changement de la position de l'objet dans la scène et donc une modification des indices visuels.
- Le τ *visuel* définit la durée que va mettre l'objet pour toucher l'auditeur ou l'auditeur pour toucher l'objet.

1.4 Combinaison des indices

Notre perception du monde est multisensorielle : de multiples indices de perception sont disponibles pour le sujet. La position d'un objet dans l'espace est le résultat de l'intégration par le cerveau des différentes informations sensorielles. Par exemple, le traitement cognitif de la disparité binoculaire est très influencé par l'angle de convergence des deux yeux. Cependant, tous les indices n'ont pas la même contribution dans cette intégration [Landy et al., 1995]. Celle-ci se base sur des indices aussi bien structurels (alignement spatial et temporel entre les indices visuels et auditifs) que cognitifs (cohérence sémantique entre le signal visuel et auditif) [Spence, 2007]. Les deux sections précédentes ont défini les indices visuels et auditifs utilisés pour localiser un objet en distance. Les sections suivantes introduisent les méthodes de combinaison des indices perçus afin de définir avec précision la position d'un objet dans l'espace.

1.4.1 Processus d'intégration

Les sujets vont combiner les différents indices afin de déterminer la position précise de l'objet visuel et/ou sonore. Un processus d'intégration a été développé par Landy et al. [1995] puis modifié par Philbeck and Loomis [1997] pour la modalité visuelle, puis repris par Zahorik et al. [2005] pour la modalité auditive. Ce processus se passe en quatre étapes, voir figure 1.7.

1. *Perception/Ajustement*

   La scène visuelle ou sonore est perçue par le sujet. Le sujet détermine les indices de profondeur. En parallèle, il y a un ajustement de la perception en fonction de différents facteurs extérieurs. Ces facteurs sont liés à la méthodologie (description verbale, triangularisation), aux connaissances du sujet sur l'objet et l'environnement, et aux mouvements du sujet (variations des indices). En effet, Andre and Rogers [2006] décrivent que deux processus d'estimation de la distance égocentrique coexistent : un premier (*ambiant*) lorsque le sujet doit estimer la distance qui le sépare d'un objet, et un deuxième (*focal*) lorsque le sujet doit activement se déplacer jusqu'à l'objet.

2. *Comparaison*

   Chaque indice produit sa propre estimation de la distance. Les estimations sont ensuite comparées deux à deux.

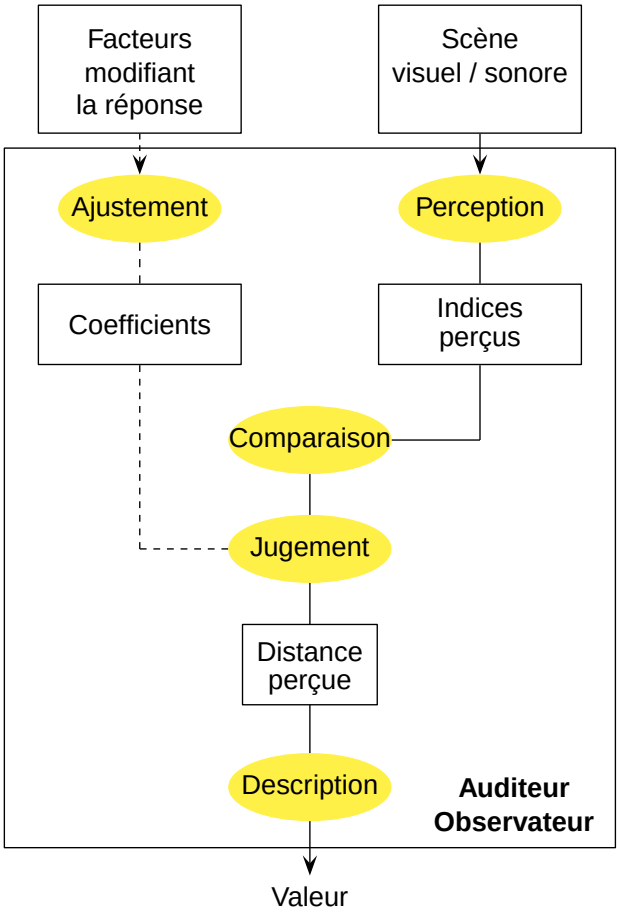


Figure 1.7 – Processus de quantification de la distance égocentrique d'un objet visuel ou sonore.

3 *Jugement*

Un coefficient est assigné à chaque estimation quantifiant le degré de fiabilité de l'indice perçu. Ces coefficients sont déterminés à partir des facteurs extérieurs et de l'influence de chaque indice. Ensuite, toutes les estimations sont intégrées en une unique estimation de la distance de l'objet visuel ou sonore (processus de *weak fusion*).

4 *Description*

La valeur finale correspond à une description par le sujet de la distance perçue en fonction de la méthodologie employée.

1.4.2 Multimodalité

Nos yeux, pris indépendamment, nous permettent d'obtenir des informations pour les deux dimensions, verticale et horizontale. Suivant l'approche empirique de la perception, cette capacité est acquise par l'expérience, l'apprentissage. Cependant, la *Gestalt* théorie considère qu'un objet est perçu dans son ensemble [Koffka, 1935]. Chaque objet de l'espace a une structure, définissant un ensemble de modalités, qui lui est propre. Ainsi, la perception multimodale d'un objet est plus que la perception de chaque modalité prise séparément [Matlin and Foley, 1997] :

> "Le tout est supérieur à la somme des parties."

Ainsi, l'ajout d'une nouvelle modalité peut influencer la perception d'un objet dans son ensemble.

Shelton and Searle [1980] ont montré que la localisation par le biais du signal sonore seul est moins précise que la localisation par le biais du signal visuel ou dans un contexte bimodal. D'après ces auteurs, la vision permet aux sujets de se positionner dans l'espace, de définir un "cadre de référence" du monde perçu. Par exemple, Nguyen et al. [2009] ont montré la dominance de la localisation visuelle sur la localisation auditive. L'intégration audio-visuelle d'indices congruents (i.e. présentation bimodale) permet une diminution de la variation inter-sujet, de l'ordre de 2,1˚pour l'écart-type en azimut. Suivant Shelton and Searle [1980] la vision améliore la localisation devant, sur le côté et derrière la tête des sujets dans le plan horizontal mais a peu d'influence dans le plan vertical. Majdak et al. [2008] ont étudié les différentes méthodes de pointage pour la localisation (tête ou main) en azimut et élévation. Bien que la vision semble améliorer les performances de localisation majoritairement en azimut, elle semble avoir un impact assez faible sur les performances de localisation en élévation.

1.4.3 Familiarité

Suivant Bronkhorst and Houtgast [1999], les auditeurs combinent deux types de sources d'information pour estimer la distance d'une source sonore : (i) des informations directement perçues et (ii) des informations indirectes correspondant à l'adaptation du sujet à la source sonore ou visuelle et à son environnement d'écoute ou d'observation, i.e. la familiarité avec la source sonore comme par exemple la voix humaine (voir étape *perception/ajustement* dans la section 1.4.1). Par exemple, Kopčo et al. [2004] ont montré que la répétition de la même source sonore à différentes distances dans le même environnement d'écoute permettait d'entendre les changements relatifs pour les indices acoustiques comme l'intensité ou le spectre.

1.4.4 Incohérences

Les indices sonores et visuels doivent rester cohérents afin de produire une source unique. Quand ces indices sont cohérents, ils conduisent à une amélioration de la perception en profondeur. Les indices sont considérés comme cohérents lorsqu'ils sont situés dans une fenêtre spatiale et temporelle d'intégration visuo-auditive de 3˚ en azimut et de 100 ms, respectivement. Lorsque les indices auditifs et visuels se situent dans cette fenêtre, ils sont toujours perçus comme coïncidents [Lewald et al., 2001]. Cependant, en cas de conflit d'indices (i.e. non coïncidents), ils conduisent à une détérioration de la perception, i.e. une diminution de la précision ou une déviation vers la position de l'indice ayant le plus d'influence sur la perception. Deux types d'incohérence peuvent être introduits entre les indices sonores et visuels, des incohérences temporelles et spatiales.

Incohérences temporelles

La fenêtre temporelle d'intégration visuo-auditive définit le point de simultanéité subjective, i.e. le décalage temporel maximum entre l'indice sonore et visuel permettant d'obtenir une sensation de simultanéité [Fujisaki et al., 2004].

Cependant, il existe une incohérence temporelle naturelle entre les indices visuels et sonores due à la différence de vitesse de propagation dans l'air de l'onde acoustique (environ 341 m/s) et de l'onde lumineuse (environ 300 000 000 m/s). Par exemple, pour un éclair lointain, l'éclair et le tonnerre ne sont pas simultanés. Bien que cet effet soit présent naturellement, les auditeurs ne semblent pas pouvoir faire le lien entre la distance cible d'un objet et le délai induit entre les indices visuels et sonores [Heron et al., 2007].

Incohérences spatiales

Lorsque les indices visuels et auditifs ne sont pas cohérents (dans l'espace), la localisation auditive est déplacée vers le stimulus visuel. Cette tendance à déplacer l'indice sonore est appelée *effet ventriloque*, ou *capture visuelle*, voir Howard and Templeton [1966] ou plus récemment Lewald et al. [2001]. L'effet ventriloque se produit principalement en azimut, mais également en distance. En effet, Gardner [1968] a montré que dans une chambre anéchoïque, un indice visuel attire les indices sonores (*proximity-image effect*), mais que cet effet est réduit en présence de réverbération [Zahorik, 2001].

ENVIRONNEMENTS VIRTUELS

La perception de la distance en milieu naturel a été étudiée de manière exhaustive. Cependant, peu d'études se sont focalisées sur la perception dans un monde virtuel. Même si une étude dans un milieu naturel a plus de validité écologique, elle ne permet pas un contrôle total de toutes les variables de l'expérimentation. Au contraire, une étude utilisant la réalité virtuelle permet une plus grande liberté dans le choix des conditions et des variables de l'expérience. Ainsi, la réalité virtuelle est utilisée dans cette étude afin de positionner plusieurs sources sonores et visuelles dans un espace en trois dimensions.

Ce chapitre détaille les techniques utilisées en réalité virtuelle pour la simulation d'indices visuels et auditifs de profondeur. Il est donc nécessaire de faire une revue de littérature exhaustive afin de sélectionner les outils nécessaires pour ce type d'application. De plus, une comparaison de la perception dans un Environnement Virtuel (EV) et un dans environnement réel est proposée pour chaque modalité.

2.1 Définition

Les applications de réalité virtuelle se caractérisent par trois facteurs essentiels, l'*autonomie* du monde virtuel, l'*interaction* entre l'humain et le monde virtuel et l'*immersion* de l'humain dans le monde virtuel. L'étude présentée dans ce rapport se focalise sur la troisième caractéristique, l'immersion. Les EV permettent de donner aux utilisateurs une sensation de présence : l'utilisateur à la sensation d'être dans l'EV plutôt que dans la pièce intégrant le simulateur. Durant de nombreuses années la réalité virtuelle était limitée au rendu visuel d'images de synthèse, si possible sur un écran stéréoscopique. De nos jours, plusieurs modalités sont utilisées comme la vision, l'audition voire le toucher grâce aux bras à retour d'effort. En effet, la réalité virtuelle permet un rendu dynamique, interactif, immersif et multimodal d'un environnement virtuel.

2.2 Interfaces visuelles

Nous avons vu dans la Section 1.3 que la distance d'une source visuelle réelle est généralement sous-estimée. Même si, Willemsen et al. [2008] et Loomis and Knapp [2003] ont montré une plus grande compression de la distance égocentrique cible dans un EV par rapport à un environnement réel, de nombreuses études ont montré une cohérence au niveau de la perception de la distance entre environnement visuel réel et virtuel [Plumert et al., 2005, Ziemer et al., 2006, Murgia and Sharkey, 2009]. Ces résultats suggèrent que la réalité virtuelle peut être utilisée pour estimer la distance égocentrique d'objets visuels.

Pour exploiter au mieux les capacités du systèmes visuel humain, la capacité des interfaces visuelles utilisées en réalité virtuelle est évaluée suivant quatre caractéristiques basées sur celles introduites par Fuchs [2003] :

- son champ de vision horizontal et vertical,
- la présence d'un rendu stéréoscopique, qui améliore la perception de la profondeur,
- le contenu visuel
- la stabilité de l'environnement visuel.

Ces quatre caractéristiques ont été adaptées . Ainsi, les applications de réalité virtuelle utilisent des interfaces visuelles spécifiques comme les visio-casques ou les *visio-cubes*. Ce type d'interfaces permet une immersion globale de l'utilisateur dans la scène visuelle.

2.2.1 Champ de vision

Afin d'augmenter la sensation d'immersion de l'observateur, les interfaces visuelles peuvent couvrir le champ de vision du système visuel humain. Pour cela, des interfaces visuelles composées de plusieurs écrans suffisamment grand sont utilisées. On appelle *visio-salle* un environnement intégrant ce type d'interface visuelle. Lorsque des vidéoprojecteurs sont utilisés, deux types de projection peuvent être utilisés : (i) la projection frontale ou (ii) la rétro-projection. Dans le second cas, l'observateur peut être proche de l'écran sans projeter son ombre sur celui-ci. Cependant, la rétro-projection nécessite une salle de projection à l'arrière de l'écran.

Une version plus poussée que la visio-salle correspond au *visio-cube*, encore appelé Cave Automatic Virtual Environment (CAVE) du nom du premier système conçu utilisant ce type d'interface. Le visio-cube propose une immersion totale du regard sur 4 (devant, gauche, derrière et droite) ou 6 (devant, gauche, derrière, droite, dessus et dessous) faces, voir figure 2.1. Il existe peu de visio-cube car ils nécessitent un espace conséquent afin de positionner les vidéoprojecteurs.

Un autre type d'interface visuelle permettant un grand champ de vision correspond aux *visio-casques*, ou Head-Mounted Display (HMD), voir figure 2.2. Cet interface à l'avantage

Figure 2.1 – Exemple de visio-cube ou Cave Automatic Virtual Environment (CAVE).

Figure 2.2 – Exemple de visio-casque ou Head-Mounted Display (HMD).

d'immerger totalement l'observateur dans la scène virtuelle. De plus, les HMD nécessitent peu d'espace car les sujets peuvent être placés dans une pièce qui n'a pas été spécialement conçue pour mettre en œuvre des applications de réalité virtuelle.

Bien que Creem-Regehr et al. [2005] aient montré que le champ de vision n'a pas d'influence sur la perception de la distance dans un environnement réel, les sujets obtiennent de meilleures performances avec une interface visuelle du type CAVE par rapport à un unique écran positionné devant l'observateur [Klein et al., 2009]. En effet, un environnement du type CAVE permet une vision périphérique. De plus, d'après Wu et al. [2004] c'est la continuité du sol entre l'environnement réel (salle de test) et l'environnement virtuel qui permet une meilleure estimation de la distance. Ainsi, pour Plumert et al. [2005], l'utilisation d'un CAVE permet d'obtenir des performances identiques (i.e. sous-estimation identique) à un environnement réel.

2.2.2 Stéréoscopie

Un rendu stéréoscopique correspond à la simulation de la profondeur à partir de deux images 2D d'un même objet prises avec deux angles légèrement différents. Ces deux angles créent une disparité, ou parallaxe, correspondant à la distance entre les deux images projetées sur un écran, pour un même point dans l'espace.

Même si la stéréoscopie permet, en théorie, une meilleure restitution de la distance et de la profondeur, la vision binoculaire permet une meilleure localisation d'un objet en champ proche uniquement, voir section 1.3.2. Ainsi, [Willemsen et al., 2008] obtiennent le même résultat avec un visio-casque en utilisant une vision binoculaire, bi-ocular (la même image sur les deux yeux) et monoculaire.

Techniques de rendu stéréoscopique

Les deux images destinées à l'œil gauche et à l'œil droit sont visualisées simultanément. Il existent différentes techniques de séparation des images gauches et droites. Celles-ci sont décrites ci-dessous.

- Différenciation colorimétrique

  Les *anaglyphes* sont deux images de couleurs différentes visualisées simultanément, habituellement rouge et cyan (i.e. deux couleurs complémentaires). L'utilisation de filtres de couleur placés sur une paire de lunettes permet à chaque œil de voir qu'une des deux images.

- Lunettes polarisantes

  Deux images en opposition de phase sont visualisées simultanément par un observateur portant des lunettes polarisantes. Il existe deux technologies pour la polarisation des images et des lunettes : (i) la polarisation linéaire utilisant deux filtres polarisants croisés à 90˚qui sont similaires aux filtres positionnées sur les vidéoprojecteurs, et (ii) la polarisation circulaire qui permet aux observateurs de garder le rendu stéréoscopique lorsqu'ils penchent la tête. Cependant, les filtres polarisants induisent une perte d'intensité lumineuse mais sont moins onéreux que les lunettes avec obturateurs.

- Lunettes actives

  Cette technologie utilise le multiplexage temporel des images dédiées à l'œil droit et à l'œil gauche. Pendant que le moniteur affiche les images à la fréquence de 120 Hz, des lunettes "actives", composées de deux écrans Liquid Crystal Display (LCD), obturent alternativement chaque œil à une fréquence de 60 Hz. Cette technique requiert une synchronisation des obturateurs avec l'affichage du moniteur, soit par câble, soit par liaison infrarouge.

- Écran auto-stéréoscopique

  Ce type d'écran permet un rendu visuel en trois dimensions sans porter de lunette. Plusieurs technologies sont utilisées : (i) l'utilisation de fines colonnes illuminées derrière l'écran à cristaux liquides ou LCD, et (ii) l'utilisation d'un réseau de lentilles demi-cylindriques. Dans les deux cas, le but étant que chaque œil puisse voir une colonne sur deux de l'écran. Cependant, ces écrans présentent l'inconvénient de diminuer la résolution des images par deux au minimum.

Sur l'ensemble des techniques de rendu stéréoscopique, on peut distinguer deux types de rendu stéréoscopique :

- La stéréoscopie *passive* : séparation par différentiation colorimétrique (anaglyphes), par lunettes polarisantes ou par réseaux lenticulaires.
- La stéréoscopie *active* : séparation par lunettes à obturateur électronique ou par réseau de colonnes illuminées.

Calibration

Avant l'utilisation d'un système stéréoscopique, il est nécessaire de calibrer la projection des deux images destinées à l'œil gauche et à l'œil droit. En effet, le système doit être calibré pour permettre une fusion des deux images par la cortex visuel, et ainsi introduire la sensation de profondeur. Cette calibration intervient au niveau du contenu à visualiser (i.e. création des deux images) et au niveau matériel (i.e. vidéoprojecteur).

La calibration du contenu à visualiser correspond à la détermination de l'angle maximum de parallaxe horizontal (de l'ordre de 1,2 à 1,5˚) et l'angle maximum de parallaxe vertical (0,3˚). Une mauvaise calibration peut entraîner une *rivalité binoculaire*, inhibant l'un des

deux yeux. De plus, il n'est pas recommandé de créer des objets "sortant de l'écran" (i.e. parallaxe négative), car ce phénomène peut empêcher la fusion des images et perturbe l'observateur. Il est intéressant de noter que suivant Yeh and Silverstein [1990], la couleur rouge semble plus facile à fusionner que la couleur blanche (limite de fusion de 6,2˚pour le rouge et 3,7˚pour le blanc).

La calibration du matériel correspond à la définition de la distance *orthostéréoscopique* entre l'observateur et le (ou les) écrans. Cette distance étant la seule susceptible de ne pas déformer les objets selon l'axe de profondeur. Il est également nécessaire de calibrer la distance inter-pupillaire des sujets. Cependant, d'après Willemsen et al. [2008] l'utilisation d'une distance inter-pupillaire individuelle donne les mêmes distances perçues qu'une distance inter-pupillaire fixe de 65 mm.

2.2.3 Contenu visuel

Le contenu visuel correspond à la résolution graphique du système de restitution (vidéo-projecteur, écran LCD) mais également à la nature de la scène visuelle.

- La scène visuelle peut être en images de synthèse ou formée à partir d'images réalistes. Cependant, d'après Willemsen and Gooch [2002], les deux types d'images produisent la même perception de profondeur. En effet, d'après les auteurs c'est le système de rendu visuel (dans ce cas un visio-casque) qui est à l'origine de la sous-estimation de la distance par rapport à un environnement réel. De plus, dans le cas où l'environnement virtuel ne représente pas un environnement réel, la quantité d'information semble avoir une influence sur la perception de la distance égocentrique. En effet, Murgia and Sharkey [2009] montrent que c'est l'indice visuel *perspective* qui semble avoir la plus grande influence dans un environnement virtuel et l'absence de cet indice entraîne une diminution de la distance égocentrique perçue.
- La scène visuelle peut soit (i) reproduire tout ou partie de l'environnement réel entourant l'observateur, ou (ii) correspondre à un environnement virtuel différent. Les visio-casques permettent de découpler entièrement l'environnement réel (i.e. salle de test) de l'environnement virtuel. Interrante et al. [2006] montrent qu'une reproduction fidèle de la salle de test permet une meilleure estimation de la distance égocentrique.

2.2.4 Stabilité de l'environnement visuel

L'immersion du regard de l'observateur dans la scène visuelle est artificiellement créé par un système de suivi de mouvements permettant de changer le point de vu sur les objets virtuels en fonction des mouvements de l'observateur. Ce changement de point de vue permet de créer une sensation de stabilité de l'environnement visuel. Ces mouvements créent un effet de *parallaxe*. Cependant, l'utilisation d'un système de suivi de mouvements implique une localisation suffisamment précise et rapide de l'observateur et une restitution visuelle fluide.

2.2.5 Évaluation des interfaces visuelles

Fuchs [2003] propose une série de tests d'évaluation des interfaces visuelles, en prenant en compte des critères physiques et perceptifs. Les critères physiques sont le chevauchement temporel des images en relief, ou *crosstalk*, qui peut introduire des images fantômes

(e.g. visualise l'image destinée à l'œil gauche au niveau de l'œil droit). Les critères perceptifs sont liés au type de visualisation :

- Monoculaire : la qualité des images (résolution de l'écran, contrastes et luminance) et la fluidité des mouvements des objets.
- Binoculaire : détermination de la limite de fusion des deux images, la qualité de perception de la profondeur.

Des critères ergonomiques, liés aux systèmes de restitution, comme le poids du visiocasque ou la fatigue visuelle due au rendu stéréoscopique peuvent être évalués. En effet, dans les systèmes de rendu stéréoscopique les yeux de l'observateur doivent accommoder au niveau l'écran. Cependant, l'observateur perçoit une sensation de profondeur pour des objets en arrière ou en avant de l'écran par le biais de la convergence des axes optiques sur les objets. Cette modification de la relation convergence-accommodation est susceptible de générer chez l'observateur un inconfort, de la fatigue visuelle voire une migraine dans le cas d'une visualisation prolongée.

Une bonne préparation du contenu à visualiser et une bonne calibration du système de rendu visuel, permet de réduire cet inconfort visuel. Par exemple, Lambooij et al. [2009] préconisent un angle maximum de parallaxe horizontal (i.e. disparité) de 1° entre les deux yeux au niveau de l'écran. Pour une accommodation à 2 m, cette disparité maximale induit une zone de confort visuelle de −1,3 à 4,8 m autour du point d'accommodation.

2.3 Interfaces auditives

Un rendu sonore spatialisé réaliste permet d'améliorer la sensation d'immersion, ou de présence, dans un EV [Zhou et al., 2004]. Par exemple, un système de restitution sonore spatialisé permet de positionner arbitrairement une source sonore à une certaine distance d'un sujet en faisant varier les indices acoustiques introduit dans la section 1.2. Ces indices sont le niveau sonore, le rapport entre champ direct et champ diffus, les indices binauraux pour les sources proches de l'auditeur (i.e. en dessous de 1 m), les indices spectraux comme l'atténuation atmosphérique qui agit comme un passe-bas pour les sources sonores lointaines, et les connaissances *a priori* de la source sonore.

Cependant, le son spatialisé est souvent sous-utilisé dans les EV au détriment d'une simulation visuelle réaliste. Les recherches expérimentales sur la perception en distance utilisant les environnements virtuels suggèrent différentes conclusions pour les interfaces auditives et celles visuelles. En effet, contrairement aux interfaces visuelles, presque toutes les études ont montré que les différences entre les scènes sonores réelles et virtuelles étaient assez faibles [Bronkhorst, 1995]. Cependant, il est difficile de les comparer car le matériel utilisé et la méthodologie diffèrent pour chaque étude. Les différences entre ces études sont :

- la nature des stimuli et de la scène sonore,
- la localisation statique ou dynamique (mouvement de la source ou de l'utilisateur),
- les dimensions utilisées (profondeur et/ou élévation et/ou azimut),
- la technique de restitution sonore spatialisée,

- la tâche demandée aux auditeurs.

Chaque caractéristique a un impact sur la perception en distance de la source sonore. Les paragraphes suivants introduisent les différentes techniques de spatialisation sonore utilisées en réalité virtuelle. Une revue de la littérature nous montre l'influence de chaque caractéristique sur la perception de la distance d'une source sonore.

2.3.1 Spatialisation sonore

Suivant Rumsey [2006], la facilité avec laquelle un auditeur peut localiser une source sonore est un important facteur de la qualité globale d'un système de restitution sonore. D'autres facteurs sont liés à la qualité de la spatialisation sonore comme :

- la précision dans la localisation,
- la restitution de la grandeur de la source sonore,
- la capacité, pour un auditeur, à se focaliser sur une source en particulier (effet cocktail-party).

Différents dispositifs de restitution sonore spatialisés sont décrits dans les paragraphes suivants. Ces dispositifs permettent de positionner une source sonore dans un espace tridimensionnel (azimut ϕ, élévation θ, distance ρ).

Approches perceptives

Les approches perceptives s'appuient sur un contrôle simple des indices de localisation interauraux décris dans la section 1.2.2. Elles reposent essentiellement sur le phénomène de source fantôme (*summing localization*), voir Blauert [1997] : un même son reproduit par deux haut-parleurs positionnés à égale distance de l'auditeur et de manière symétrique par rapport au plan médian donne l'impression que la source sonore se situe entre les deux haut-parleurs, i.e. au niveau du plan médian. Suivant le nombre de haut-parleurs employés, il existe différents systèmes de restitution :

- La stéréophonie se base sur deux haut-parleurs positionnés de chaque côté de l'auditeur, e.g. $\pm 30^{\circ}$par rapport au plan médian.
- Le home-cinéma (systèmes 5.1 et 7.1), voir [ITU–R Rec. BS.775–2], utilise plus de haut-parleurs ce qui permet d'accroître l'enveloppement de l'auditeur (équivalent au champ de vision pour les interfaces visuelles).

Cette technique peut-être généralisée pour le rendu de scènes sonores en trois dimensions, voir Pulkki [1997]. En effet, la stéréophonie et les systèmes home-cinéma permettent de reproduire une source sonore dans le plan horizontal uniquement. La technique Vector Base Amplitude Panning (VBAP) permet de positionner plusieurs sources sonores autour de l'auditeur en appliquant des gains à chaque haut-parleur utilisé. En théorie, le nombre de haut-parleurs est illimité mais il est possible d'obtenir une restitution cohérente avec peu de haut-parleurs (e.g. 4–6).

Cependant, les systèmes utilisant une approche perceptive sont du type mono-utilisateur. En effet, l'auditeur doit être positionné au centre du dispositif de restitution (*sweet spot*) pour que la source sonore soit localisée correctement.

Technique binaurale

Dans le cas d'une écoute individuelle, la technique binaurale peut-être utilisée. Elle permet une reproduction fidèle d'une scène sonore en trois dimensions. Cette technique encode les indices de perception binauraux introduits dans la section 1.2.2 (i.e. l'ITD, l'ILD et les indices spectraux) par le biais des HRTF. Bronkhorst [1995] détaille le processus de mesure des filtres HRTF, H_G et H_D. Cette mesure est assez complexe :

- Un haut-parleur produit un bruit large-bande dans chaque direction (azimut, élévation et distance) qui est enregistré par une sonde (microphone) située à l'entrée du canal auditif du sujet, voir figure 2.3. Le haut-parleur est généralement situé à une distance supérieure à 1 m de l'auditeur afin de réduire les résonances qui apparaissent en champ proche, voir 1.2.2.
- Il faut ensuite éliminer l'influence :
  - du bruit large-bande,
  - de la salle d'enregistrement (réflexions), ou d'enregistrer dans une chambre "anéchoïque" (i.e. non réverbérante),
  - du haut-parleur utilisé lors de la mesure.
- De plus, les mesures sont faites à l'entrée du canal auditif du sujet, ou Drum Reference Point (DRP). Cependant le signal sera reproduit par le biais d'un casque, au niveau de l'Ear Reference Point (ERP), i.e. la position du haut-parleur. Il est donc nécessaire d'appliquer un filtre correspondant au chemin DRP vers le point de restitution, e.g. ERP.

Il en résulte que les fonctions de transfert, $T(\theta,\phi)_G$ et $T(\theta,\phi)_D$ (une par oreille), qui sont multiplié avec le signal sec (i.e. monophonique et anéchoïque) à spatialiser, sont définis par [Bronkhorst, 1995] (ici pour l'oreille gauche) :

$$T(\theta,\phi)_G = \frac{H(\theta,\phi)_G \cdot P}{F \cdot S}\ , \tag{2.1}$$

où $H(\theta,\phi)_G$ est le filtre HRTF pour l'oreille gauche, l'azimut ϕ et l'élévation θ, P correspond au spectre du bruit large-bande, F est la fonction de transfert du système de reproduction utilisé lors de la restitution (haut-parleur), et S correspond au spectre du signal large-bande diffusé par le haut-parleur lors de la mesure.

Le rendu *binaural* sur casque est une technique de rendu sonore assez facile à mettre en œuvre. Elle ne requiert qu'un dispositif limité et reproduit le champ sonore exact aux oreilles de l'auditeur. Cependant, son utilisation dans une application de réalité virtuelle introduit certaines difficultés et montre les limitations de cette technique :

- Il faut mesurer les filtres HRTF pour chaque sujet. Cependant, même avec une utilisation de filtres individualisés (i.e. de chaque utilisateur) il existe un manque de précision en localisation. D'après Bronkhorst [1995], ce manque de précision semble être liés aux erreurs du rendu sonore au delà de 7 000 Hz. Ces erreurs proviennent des erreurs dans le calcul des HRTF. Or, Zahorik [2002b] a montré que l'utilisation de filtres HRTF générique (e.g. ceux d'une tête artificielle) donnait la même perception en distance que des filtres HRTF individualisés.

Figure 2.3 – Ensemble de haut-parleurs utilisés pour la mesure de HRTF. Le sujet est positionné sur la chaise au milieu du cercle de haut-parleur.

- L'existence de cônes de confusion, i.e. une inversion possible entre l'avant et l'arrière (ou arrière-avant) et entre le dessus et le dessous (ou dessous-dessus). Ce phénomène correspond à la localisation du stimulus sonore à l'opposée de le source cible par rapport à l'axe interaural. Pour éviter ce problème, Begault et al. [2001] propose l'utilisation d'un système de suivi de mouvements et d'un stimulus sonore d'une durée supérieure à 3 s. En effet, les auditeurs peuvent corriger ces erreurs de localisation par le biais de rotation de la tête.
- La source sonore peut-être localisée dans la tête de l'auditeur dans le cas où la source n'est pas externalisée. L'introduction d'un effet de salle (réverbération) permet d'augmenter le phénomène d'externalisation [Begault, 1992].
- Les filtres HRTF définissent un point précis dans l'espace. Dans les application de réalité virtuelle, les sources sonores peuvent être en mouvement dans la scène sonore. Cela implique une interpolation en continu (voire en temps-réel dans le cas d'un système interactif) des filtres HRTF.

Technique transaurale

La technique du *binaural sur haut-parleurs*, souvent appelée *transaurale*, permet de simuler un rendu binaural à partir de deux haut-parleurs positionnés de chaque côté de l'auditeur, voir Kirkeby et al. [1997]. Cependant, elle requiert une phase de calibration où l'effet croisé de chaque haut-parleur est annulé. Les réflexions de la salle de test peuvent compliquer cette phase de calibration. Ainsi, il est recommandé d'utiliser la technique transaurale dans un environnement proche d'une salle anéchoïque.

(a) Vue de devant (b) Vue de derrière

Figure 2.4 – Système de réalité virtuelle SMART-I^2. La vue de derrière montre les *Multi-Actuator Panel* (MAP).

Holophonie

Le *binaural* est utilisé dans des dispositifs de réalité virtuelle ou de réalité augmentée, comme les CAVEs, surtout dans le cas ou des bruits parasites sont produits par les interfaces visuelles (i.e. vidéoprojecteurs) ou lorsque l'environnement est réverbérant. Cependant, l'utilisation de casques peut réduire le niveau d'immersion du sujet et aussi l'interactivité ou les communications dans les systèmes multiple-utilisateurs. L'holophonie est un système de rendu sonore spatialisé applicable aux situations multiple-utilisateurs.

L'holophonie, ou Wave Field Synthesis (WFS), se base sur un formalisme mathématique décrivant les équations de propagation du son, voir Berkhout et al. [1993]. Cette technique permet de reproduire le champ sonore autour de l'auditeur. Cependant cette technique requiert un nombre conséquent de haut-parleurs (surtout pour une reproduction tridimensionnelle, incluant l'élévation). De plus, cette technique ne reconstruit que partiellement le signal auditif. Par exemple, le nombre limité de haut-parleurs induit un échantillonnage spatial et donc une fréquence maximale de reproduction (cf. critère de *Shannon*), e.g. 850 Hz pour un échantillonnage de 0,2 m [Spors and Ahrens, 2009].

Le système de réalité virtuelle appelé SMART-I^2, situé au Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur (LIMSI), voir Rébillat et al. [2008], utilise la technique WFS. Ce système est présenté dans la figure 2.3.1. Cependant, les haut-parleurs sont remplacés ici par des *Multi-Actuator Panel* (MAP) positionnés derrière l'écran utilisé pour la visualisation, voir figure 2.4(b).

Décomposition en harmoniques sphériques

L'holophonie décrit un système de reproduction sonore qui permet de positionner des sources sonores virtuelles. Les approches utilisant la décomposition en harmoniques sphériques, comme Higher Order Ambisonics (HOA), permettent un encodage du champ sonore, voir Daniel et al. [2003]. Cette technique est donc utilisée au niveau de l'enregistrement, par le biais de microphones *soundfield* voir figure 2.5. Le signal enregistré peut ensuite être restitué sur quasiment tous les dispositifs. La technique ambisonique (harmoniques d'ordres 0 et 1) induit certaines contraintes pour l'utilisateur comme une position fixe

Figure 2.5 – Microphone *Soundfield*© utilisé pour enregistrer en utilisant la technique de décomposition en harmoniques sphériques du champ sonore.

au centre de l'espace de diffusion, *sweet spot*. Cependant, une augmentation du nombre d'harmoniques sphériques restituées permet d'agrandir la taille du sweet-spot.

Différentes techniques de spatialisation peuvent être combinées. Par exemple, Noisternig et al. [2003] proposa une approche ambisonique virtuelle combinée à un système de restitution binaural.

2.3.2 Stimulus et scène sonore

Plusieurs types de stimuli peuvent être utilisés dans les applications de réalité virtuelle comme la parole, la musique et un ensemble de sons d'ambiance. Suivant Zahorik [2002a], la perception en distance d'une source sonore n'est pas influencée par son type. En effet, cet auteur ne trouve aucune différence significative entre un signal de parole et un bruit pulsé. Cependant, Lokki et al. [2000] montrent qu'un bruit rose permet une meilleure navigation dans un environnement virtuel que des sons provenant d'instruments de musique.

2.3.3 Simulation de salle d'écoute

Dans le cas d'une diffusion sur un système de haut-parleurs distants (WFS, VBAP, ...) la salle de test peut avoir une influence sur l'onde acoustique perçue par les sujets. Dans ce cas, il est nécessaire d'annuler l'impact des réflexions sur les parois de la salle de test du signal sonore reproduit par les haut-parleurs. La technique binaurale permet de découpler le signal sonore perçu de la salle de test. Dans ce cas, il est possible de simuler les effets de salles, ou *auralisation*. L'introduction d'une réverbération permet d'externaliser la position de la source sonore et également de produire une environnement sonore virtuel réaliste [Durlach et al., 1992] voire d'augmenter la distance égocentrique perçue [Begault, 1992].

Il est possible de simuler l'acoustique d'une salle d'écoute par le biais de sa réponse impulsionnelle ou Room Impulse Response (RIR). Celle-ci peut être calculée de différentes manières :

- *Enregistrements* dans une salle réelle : par exemple Jeub et al. [2009] ont enregistré avec une tête artificielle une série de RIR dans différentes salles et pour différentes distances.

- *Simulations* par le biais de lois géométriques et/ou statistiques : par exemple, l'algorithme *Roomsim* de Campbell et al. [2005], calcule des RIR binaurales à partir d'une technique de lancé de rayons.

[Noisternig et al., 2003] proposa un calcul de simulation d'une salle d'écoute se déroulant en deux phases. Dans un premier temps, les réflexions précoces aux ordres 0 et 1 sont calculées à partir d'une modélisation géométrique de la salle d'écoute. Les réflexions tardives sont implémentés à partir de réseaux de délais. Shinn-Cunningham [2000a] et Kopčo et al. [2007] montrent que les indices binauraux d'une réverbération n'ont que peu d'impact sur la perception en distance d'une source sonore. Ainsi, la simulation de la réverbération peut être simplifiée en utilisant la même RIR pour les deux oreilles, gauche et droite.

Begault et al. [2001] a montré que l'ajout d'une réverbération permet de réduire l'erreur de localisation en azimut mais cette erreur augmente en élévation. De plus, suivant Kopčo et al. [2007], la réverbération permet une meilleure précision des sujets dans l'estimation de la distance. Cependant, pour un rendu dynamique, Lokki et al. [2000] montrent que l'introduction de réverbération (effet de salle) semble compliquer la navigation dans un environnement virtuel : la réverbération introduit un étalement de la source sonore qui rend la localisation moins précise.

2.3.4 Stabilité de l'environnement sonore

Les techniques de restitution sonore spatialisée utilisant des haut-parleurs distant comme la stéréophonie ou l'holophonie permettent de garder une scène sonore stable lors des mouvements de l'auditeur, i.e. la scène sonore reste fixe lorsque l'auditeur tourne la tête. Cela est vrai uniquement lorsque l'auditeur reste dans une zone d'écoute appelée *sweet spot*. Cependant, pour une reproduction au casque, la technique binaurale peut être couplée avec un système de suivi de mouvements de la tête qui permet à la source sonore virtuelle de rester à une position fixe dans l'environnement et donc variable relativement à l'orientation de l'utilisateur. Cela impose un rendu dynamique, i.e. une interpolation en temps-réel des filtres HRTF. L'utilisation d'un système de suivi de mouvements de la tête couplé à un rendu dynamique, voir Begault et al. [2001], diminue l'inversion avant/arrière et améliore la localisation en azimut et en élévation des sources sonores virtuelles. Cependant, d'après Begault et al. [2001], l'utilisation d'un traqueur de tête et d'un rendu dynamique, améliore la localisation uniquement pour les filtres HRTF génériques et non pour les filtres HRTF individualisés.

2.4 Cohabitation des systèmes de rendus

Bien que considérés comme temps-réels, les systèmes de rendu visuel et sonore ont une latence globale ce qui peut créer une dé-synchronisation dans le rendu sonore et visuel. Cette dé-synchronisation peut perturber l'interaction entre l'utilisateur et l'application. Il est donc nécessaire de synchroniser les différents systèmes de restitution. Des outils de communication entre les différents logiciels de rendu sonore et visuel sont utilisés. Pour cela des *sockets*, utilisant par exemple de protocole UDP, permettent d'établir une connexion entre les systèmes.

De plus, la latence globale de chaque système de rendu peut avoir un impact sur la performance des sujets. Il est donc nécessaire de réduire la latence globale du système dans le cas où :

- le sujet peut modifier l'environnement virtuel (systèmes interactifs),
- le système de restitution intègre un suivi des mouvements de l'auditeur (*tracking*).

Certaines études ont montré que les sujets étaient peu sensibles à la latence du système de rendu sonore (pour des latences inférieures à 250 ms). Cependant, les sujets semblent beaucoup plus sensibles à la latence pour la modalité visuelle : seuil proche de 15 ms [Ellis et al., 2004].

PROTOCOLE EXPÉRIMENTAL

L'immersion d'un humain dans un Environnement Virtuel (EV) passe par le rendu de plusieurs modalités comme la vision, l'audition ou le toucher. Durant les deux dernières décennies, la majeure partie des applications se sont focalisées sur la restitution visuelle. En effet, les technologies de rendu visuel réaliste comme la stéréoscopie, ont concentré les efforts de recherche. Depuis peu, l'utilisation de technologies de rendu sonore a permis la restitution multimodale d'objet dans un EV. Cependant, il existe peu d'études sur la cohérence des indices visuels et sonores pour des objets virtuels. Cette étude se focalise sur la perception en distance d'objets virtuels. L'étude décrite dans ce chapitre cherche à répondre aux trois questions suivantes :

- *La perception bimodale d'un objet permet-elle une meilleure localisation en distance d'un objet dans l'espace ?*
- *Les systèmes de restitution visuelle et sonore utilisés en réalité virtuelle permettent-ils d'obtenir un rendu naturel de la distance ?*
- *Existe-t-il un espace de cohérence entre les indices visuels et sonores provenant d'un même objet ?*

Pour répondre à ces trois questions, nous allons utiliser un EV multimodal. Cependant, cela implique une restitution fiable de la distance. Le chapitre suivant décrit les systèmes de restitution visuelle et sonore utilisés dans cette étude.

3.1 Matériel

Cette étude se focalise sur la perception en distance d'objets visuels et sonores. Pour cela un même objet est placé à différentes distances d'un sujet humain. Les différentes conditions sont créées suivant un protocole expérimental décrit dans ce chapitre.

3.1.1 Conditions

Les conditions ont été choisies suivant quatre variables expérimentales :

- la modalité de présentation (visuelle, auditive ou bimodale),
- la distance de la source (les distances étudiées sont inclues dans l'espace d'action du sujet : 2, 3, 5, 10 et 20 m),
- les caractéristiques du rendu sonore (temps de réverbération T_{60}),
- la quantité d'information visuelle (nombre et disposition dans l'espace des objets visuels).

De plus, l'influence de la cohérence entre les indices visuels et sonores est étudiée par le biais de huit conditions avec des indices visuels et sonores incohérents. Pour ces conditions, l'indice sonore est déplacé vers l'avant ou vers l'arrière par rapport à l'indice visuel.

Les tableaux 3.1, 3.2, 3.3 et 3.4 listent les 48 conditions évaluées par les sujets au cours du test de perception. Ces conditions peuvent être regroupées en quatre groupes :

1. modalité auditive seule (conditions 1–10),
2. modalité visuelle seule (conditions 11–20),
3. bimodalité, auditive et visuelle, avec indices cohérents (conditions 21–40), et
4. bimodalité, auditive et visuelle, et disparité entre les indices auditifs et visuels (conditions 41-48).

Dans le tableau 3.4, la distance (2e colonne) correspond à la position de l'indice visuel. L'offset (3e colonne) détermine la distance entre la position de l'indice visuel et l'indice sonore.

Tableau 3.1 – Liste des conditions auditives évaluées au cours du test de perception.

Condition	Distance (m)	Environnement sonore
1	2	mat
2	3	mat
3	5	mat
4	10	mat
5	20	mat
6	2	réverbérant
7	3	réverbérant
8	5	réverbérant
9	10	réverbérant
10	20	réverbérant

Tableau 3.2 – Liste des conditions visuelles évaluées au cours du test de perception.

Condition	Distance (m)	Environnement visuel
11	2	Pauvre
12	3	Pauvre
13	5	Pauvre
14	10	Pauvre
15	20	Pauvre
16	2	Riche
17	3	Riche
18	5	Riche
19	10	Riche
20	20	Riche

3.1.2 Environnement et stimuli visuels

Le but de l'expérience étant d'étudier un processus cognitif, les stimuli sont les plus minimalistes possibles. D'après Yarbus [1967], il faut au minimum 200 ms pour localiser

Tableau 3.3 – Liste des conditions bimodales évaluées au cours du test de perception.

Condition	Distance (m)	Environnement sonore	Environnement visuel
21	2	mat	Pauvre
22	3	mat	Pauvre
23	5	mat	Pauvre
24	10	mat	Pauvre
25	20	mat	Pauvre
26	2	réverbérant	Riche
27	3	réverbérant	Riche
28	5	réverbérant	Riche
29	10	réverbérant	Riche
30	20	réverbérant	Riche
31	2	mat	Riche
32	3	mat	Riche
33	5	mat	Riche
34	10	mat	Riche
35	20	mat	Riche
36	2	réverbérant	Pauvre
37	3	réverbérant	Pauvre
38	5	réverbérant	Pauvre
39	10	réverbérant	Pauvre
40	20	réverbérant	Pauvre

visuellement un objet. Les stimuli étant des scènes virtuelles de 8 secondes, les sujets ont assez de temps pour localiser le stimulus visuel et/ou sonore.

Stimulus visuel

La source visuelle correspond à un haut-parleur virtuel de $40 \times 60\,\text{cm}^2$, de couleur bleue, la face avant du haut-parleur (dôme) tournée vers le sujet. Le centre du dôme du haut-parleur est positionné à la hauteur des yeux de l'observateur. Un pied est positionné en dessous du haut-parleur permettant de créer un contact entre le haut-parleur et le sol.

Environnement visuel

La source visuelle est positionnée dans un environnement virtuel qui correspond à une extension de la salle de test réelle à travers l'écran. La couleur et les dimensions des murs, du plafond et du sol de la salle virtuelle ont été choisies pour simuler la salle réelle. Ainsi, l'ensemble salle réelle et extension virtuelle doit être perçu comme une seule et unique salle. La salle réelle et son extension ont une hauteur de 2,60 m est une largeur de 5,51 m. La salle réelle a une profondeur de 7,08 m et son extension virtuelle a une profondeur de 25 m afin de garder un espace derrière le stimulus visuel lorsque celui-ci est placé à 20 m du sujet. La figure 3.1 montre le positionnement du sujet et des stimuli dans la salle réelle et virtuelle. La scène visuelle est générée par la librairie Atelier de Réalité

Tableau 3.4 – Liste des conditions incohérentes évaluées au cours du test de perception..

Condition	Distance (m)	Offset (m)	Environnement sonore	Environnement visuel
41	2	−1	mat	Riche
42	5	+15	mat	Riche
43	5	−4	mat	Riche
44	5	+5	mat	Riche
45	10	+10	mat	Riche
46	10	−7	mat	Riche
47	10	+5	mat	Riche
48	20	−15	mat	Riche

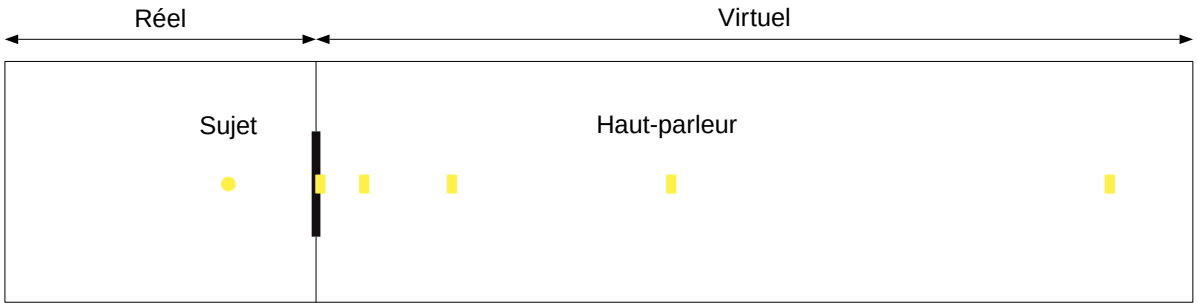


Figure 3.1 – Dimension de la salle de test et de son extension virtuelle. Les points correspondent aux positions du sujet (cercle) et des stimuli visuels (rectangles).

Virtuelle (ARéVi) [CERV, 31 March 2011], qui utilise l'interface de programmation Open Graphics Library (OpenGL). Deux environnements visuels sont simulés dans cette étude :

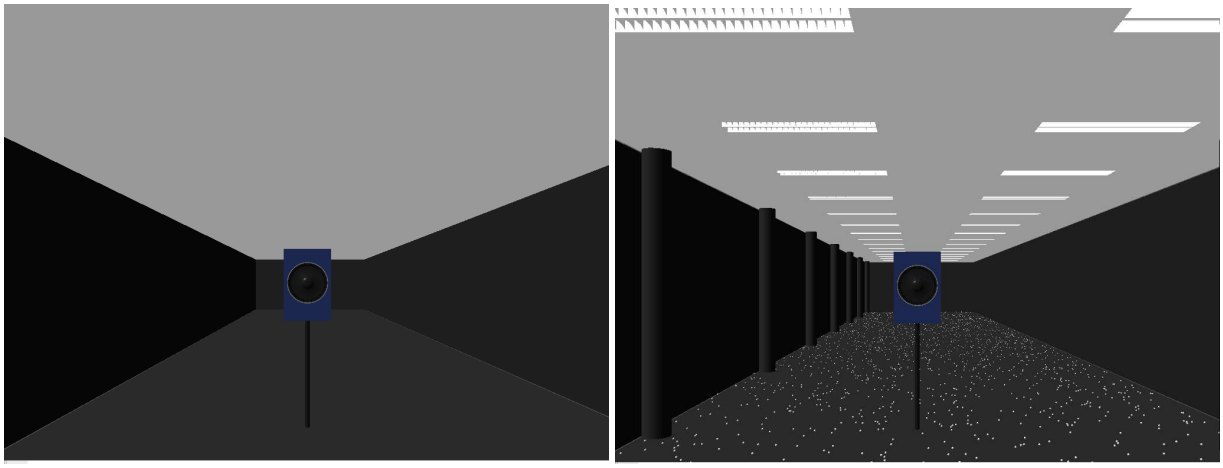
- Un environnement pauvre en indices visuels comprenant les murs, le plafond et le sol, voir figure 3.2(a). Ces indices visuels apportent principalement des information au niveau de la perspective.
- Un environnement riche en indices visuels qui en plus des indices précédent inclut une texture appliquée au sol, des colonnes et des néons, voir figure 3.2(b). Cet environnement apporte, en plus de la perspective, des ancres visuelles.

Une source lumineuse réelle est placée dans la salle de test afin d'éclairer l'interface de réponse au test utilisée par les sujets, voir figure 3.1.

Le système de restitution visuel correspond à un écran de 2,4 × 1,8 m combiné à deux vidéo-projecteurs *Barco* d'une résolution de 1280 × 1027 pixels, et d'une fréquence de restitution de 45 Hz. Les vidéo-projecteurs sont placés dans une salle annexe et équipés de filtres à polarisation circulaire (technologie stéréoscopie passive) permettant une restitution en trois dimensions de le scène visuelle. Cependant chaque participant doit porter une paire de lunettes polarisantes.

Calibration

Les sujets sont assis sur une chaise placée à 2 m de distance et en face du milieu de l'écran, voir figure 3.3. Cette position permet un champ de vision de ±31°. Une calibration



(a) Environnement pauvre en indices visuels (b) Environnement riche en indices visuels

Figure 3.2 – Environnement visuel de l'expérience.

du système de restitution visuelle nous a permis de fixer la distance orthostéréoscopique au niveau du sujet. Ainsi, à la position du sujet, l'extension virtuelle de la salle réelle respecte les proportions de la scène réelle (largeur, hauteur, profondeur). De plus, le système a été calibré pour une distance interpupillaire de 65 mm.

3.1.3 Environnement et stimuli sonores

Les deux scènes visuelles reproduisent une grande quantité d'indices visuels de profondeur comme la perspective et le gradient de texture. Afin de reproduire des indices sonores cohérents avec les indices visuels, le système de rendu sonore doit permettre une reproduction naturelle de la distance. Cela implique une variation des caractéristiques de l'environnement et du stimulus sonore comme l'intensité, le contenu spectral et le rapport entre le champ direct et le champ réverbéré. Cependant, l'influence de l'azimut et de la latéralisation sur ces indices sonores ne sera pas étudié. En effet le stimulus sonore est placé dans le plan médian et interaural.

Stimulus sonore

D'après Lokki et al. [2000], un bruit rose permet une navigation dans un espace virtuel plus rapide qu'avec un instrument de musique. Cependant, une réverbération est difficilement perçue avec un bruit rose. Un bruit rose peut donc être utilisé pour une localisation directionnelle mais pas pour une estimation en distance. Plusieurs études ont montré que la parole pouvait être utilisée comme stimulus sonore, e.g. Gardner [1969], même si la parole n'est pas le stimulus le plus "agréable" pour les sujets [Tran et al., 2000]. Par conséquent, la source sonore employée dans cette étude correspond à un signal de parole composé de deux phrases :

- la première phrase est prononcée par un homme : *Le camp d'été s'est passé au bord du fleuve*,
- la deuxième phrase est prononcée par une femme : *La voiture s'est arrêtée au feu rouge*.

De plus, la parole humaine est principalement composée de fréquence entre 50 et 7000 Hz [Deng and O'Shaughnessy, 2003]. Or, d'après Bronkhorst [1995], l'utilisation de fréquences

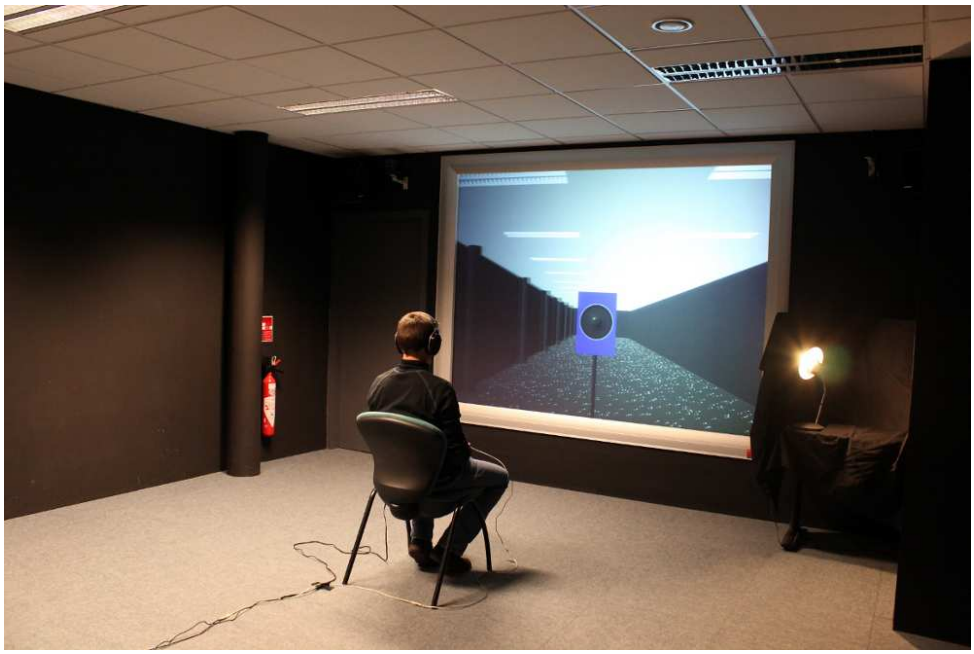


Figure 3.3 – Disposition du sujet dans la salle de test.

supérieures à 7000 Hz peut introduire des effets indésirables pour une bonne restitution binaurale de sources sonores virtuelles.

Environnement sonore

Le stimulus sonore est ensuite spatialisé par le système de rendu sonore *binaural* puis reproduit au casque. Cette méthode, bien que plus intrusive que les systèmes du type VBAP ou WFS, permet un meilleur contrôle du rendu sonore, en évitant les effets de la salle d'écoute. Le signal source est artificiellement positionné aux différentes distances listées dans les tableaux 3.1, 3.2, 3.3 et 3.4 par le biais de réponses impulsionnelles binaurales de salle ou Binaural Room Impulse Response (BRIR) en anglais. Cette réponse impulsionnelle binaurale inclut l'effet de la salle d'écoute (réverbération) et l'effet de la tête de l'auditeur (HRTF). Cependant, ces effets sont dans un premier temps enregistrés avec une tête artificielle ou simulés par un logiciel puis introduits dans le signal source par le biais d'une convolution.

Le Centre Européen de Réalité Virtuelle (CERV) ne disposant pas d'un système de mesure de filtres HRTF, des HRTF génériques (non individualisés) ont été utilisés. Zahorik [2002b] a montré que l'utilisation de filtres HRTF génériques ne réduisait pas la précision de la perception en distance. Par conséquent les BRIR utilisées on été créées à partir d'HRTF de mannequin Knowles Electronics Manikin for Acoustic Research (KEMAR) [G.R.A.S.©, Retrieved 10 June 2011]. Les BRIR utilisées sont composées de deux parties :

- Les réflexions précoces (jusqu'au deuxième ordre) sont simulées par le script Matlab "Roomsim" [Campbell et al., 2005]. Les dimensions de la salle simulées correspondent à celles de la salle de test, soit la combinaison de la salle réelle et son extension virtuelle, voir figure 3.1. Cependant, afin de réduire les effets d'échos flottants, la profondeur de la salle simulée a été réduite de 5 m, soit une salle ayant les

dimensions suivantes : 27,1 $\times$ 5,5 $\times$ 2,60 m^3. Afin de simuler deux réverbérations différentes, deux jeux de coefficients α ont été appliqués pour définir l'absorption des murs de la salle d'écoute, voir tableaux 3.5 et 3.6 [1]. Ces coefficients correspondent à des valeurs standards de matériaux utilisés dans la construction. Ces coefficients ont été choisis afin de produire deux temps de réverbération donnant l'impression d'une salle mate (T_{60} = 370 ms) ou réverbérante (T_{60} = 860 ms). Enfin, un modèle d'absorption de l'air a été utilisé (inclus dans le script "Roomsim").

- La partie diffuse de la réverbération provient d'une base données de BRIR réelles [Jeub et al., 2009]. Cette base de données a été utilisée pour simuler une réverbération réaliste, i.e. au niveau de la cross-corrélation interaurale.

Les deux parties des BRIR ont été générées à une fréquence d'échantillonnage de 44100 Hz et combinées afin de produire une réverbération réaliste. Un temps d'attaque et un déclin de 50 ms a été appliqué à chaque signal sonore. Le niveau d'écoute a été calibré pour obtenir 63 dB$_{SPL}$ au niveau de chaque oreille de l'auditeur, pour une source sonore positionnée à 2 m de l'auditeur. En effet, ITU–T Handbook on Telephonometry propose d'utiliser un niveau de référence de 69 dB$_{SPL}$ au niveau de chaque oreille de l'auditeur pour un locuteur positionné à 1 m. Pour l'ensemble des conditions des tableaux 3.1, 3.2, 3.3 et 3.4 le niveau d'écoute se situe entre 63 et 58,7 dB$_{SPL}$. Les signaux sonores sont envoyés à une carte son *Lexicon Alpha* puis diffusés par un casque *Sennheiser HD 650*.

Tableau 3.5 – Coefficients d'absorption α des murs de la salle d'écoute pour le temps de réverbération faible (T_{60} = 370 ms).

	Fréquence (Hz)					
Mur	125	250	500	1000	2000	4000
Derrière	0.14	0.35	0.55	0.72	0.70	0.65
Devant	0.14	0.35	0.55	0.72	0.70	0.65
Droit	0.14	0.35	0.55	0.72	0.70	0.65
Gauche	0.14	0.35	0.55	0.72	0.70	0.65
Sol	0.08	0.24	0.57	0.69	0.71	0.73
Plafond	0.70	0.66	0.72	0.92	0.88	0.75

3.2 Méthode

D'après Loomis et al. [1998], une action dirigée comme la triangularisation permet d'obtenir de meilleure estimation de la distance et également de réduire les différences inter-sujets. Cependant, le but de cette étude est d'étudier la perception en distance d'objets statiques. Une tâche d'estimation directe de la distance a donc été utilisée dans cette étude. L'expérience se déroule en deux sessions :

1. Durant la première session les sujets ont évalué les conditions correspondant à la modalité auditive (conditions 1 à 10) et à la modalité visuelle (conditions 11 à 20). Les conditions sont évaluées par blocs (bloc auditif et bloc visuel). Afin d'étudier si

1. La disposition des murs se réfère à l'orientation de l'auditeur dans la salle.

Tableau 3.6 – Coefficients d'absorption α des murs de la salle d'écoute pour le temps de réverbération long (T_{60} = 860 ms).

	Fréquence (Hz)					
Mur	125	250	500	1000	2000	4000
Derrière	0.36	0.44	0.31	0.29	0.39	0.25
Devant	0.36	0.44	0.31	0.29	0.39	0.25
Droit	0.36	0.44	0.31	0.29	0.39	0.25
Gauche	0.36	0.44	0.31	0.29	0.39	0.25
Sol	0.08	0.24	0.57	0.69	0.71	0.73
Plafond	0.70	0.66	0.72	0.92	0.88	0.75

la visualisation du stimulus visuel permet un apprentissage des distances et ainsi influence la perception pour la modalité auditive seule, la moitié des sujets a commencé avec le bloc auditif et l'autre moitié des sujets avec le bloc visuel. Cette première session dure 25 minutes.

2. Durant la deuxième session les sujets ont évalué les conditions bimodales (conditions 21 à 48) en trois blocs. Les deux premiers blocs présentaient les conditions avec indices visuels et sonores cohérents (conditions 21 à 40), puis les sujets ont évalué les conditions avec indices incohérents (conditions 41 à 48) dans le dernier bloc. Cette deuxième session, plus longue, dure 40 minutes.

Une pause obligatoire de 4 minutes sépare chaque bloc. Les deux sessions sont séparées en temps par 36 heures au minimum. Dans chaque bloc, les stimuli sont présentés de façon aléatoire avec une liste de lecture par sujet pour éviter un effet de l'ordre de lecture. Chaque stimulus est évalué 4 fois afin d'obtenir des réponses stables pour chaque condition.

3.2.1 Description de la tâche

Chaque essai se déroule en deux étapes :

1. le stimulus visuel et/ou sonore de 8 secondes est présenté au sujet,
2. à la fin de la présentation, il est demandé au sujet de reporter son estimation de la distance égocentrique de l'objet en utilisant un pavé numérique (la valeur rentrée s'affiche à l'écran). Le sujet doit ensuite valider cette estimation en tapant sur entrée. Le temps de réponse est limité à 12 s afin d'obtenir une réponse spontanée. Le stimulus suivant est joué automatiquement.

3.2.2 Déroulement du test

Au début de l'expérience, la hauteur du regard (i.e. hauteur des yeux) a été mesurée pour calibrer le système de rendu stéréoscopique. Cette position correspond au point d'origine pour le calcul de la hauteur du stimulus visuel. Durant la suite du test, il a été demandé aux participants d'éviter de bouger et de tourner leur tête. Ensuite, un texte explicatif a été fourni à tous les sujets, voir Annexe A. Après lecture des instructions, l'expérimentateur décrit verbalement la tâche à effectuer puis le sujet lance l'application. Celle-ci commence

par une phase d'entraînement qui permet aux sujets de se familiariser avec l'application. Pour la première session, cette phase d'entraînement incluait 8 essais, et pour la deuxième session, elle incluait 4 essais.

3.2.3 Participants

Au total, 24 sujets ont participé à l'expérience (2 femmes et 22 hommes). Tous les sujets sont des membres du CERV (enseignants-chercheurs, stagiaires, doctorants, ...). Tous sont considérés comme naïfs vis-à-vis de l'expérience mais la plupart des sujets ont des connaissances en réalité virtuelle. Aucun sujet n'a rapporté avoir de problème d'audition et tous les sujets ont une vue normale ou corrigée.

RÉSULTATS

Ce chapitre présente et analyse les résultats de l'expérience décrite dans le chapitre 3. D'abord, la section 4.1.1 revient sur les biais et les limites des résultats. Les caractéristiques générales des résultats sont présentées dans la section 4.1 puis les résultats sont présentés en deux parties : la section 4.2 présente les résultats des conditions avec indices visuels et auditifs cohérents, puis la section 4.3 présente les résultats pour les huit conditions avec indices incohérents.

4.1 Caractéristiques générales

Sur les 4608 essais évalués, 4593 essais ont été analysés. Les sujets n'ont pas réussi à estimer la distance de l'objet dans les 12 secondes pour 15 essais.

4.1.1 Biais et limites

La salle de test (appelée salle noire) située au CERV n'a pas été construite pour faire des expériences auditives. Durant les tests, plusieurs bruits étaient audibles par les auditeurs comme l'expulsion des fumées de la chaudière par le conduit situé au milieu de la salle de test. De plus, la salle n'étant pas isolée acoustiquement du reste du bâtiment, les mouvements de personnes à côté de la salle de test et les fermetures/ouvertures des portes étaient audibles par les sujets.

4.1.2 Mesure de précision

La procédure d'analyse statistique doit quantifier la performance des sujets pour chaque conditions. Ce chapitre se base sur la distance perçue, ρ_i. Elle correspond à la moyenne arithmétique de tous les jugements pour une condition donnée :

$$\rho_i = \frac{1}{M} \sum d_{i,k,l} \tag{4.1}$$

où $i = 1 \dots I$ est la condition, $k = 1 \dots K$ est le sujet, $l = 1 \dots L$ est le nombre de répétition, et $M = K \cdot L$ le nombre de jugements individuels. De plus, un écart-type, σ_i, et un intervalle de confiance à 95%, CI_i, sont calculés pour chaque condition. Pour des valeurs ayant une distribution normale, l'intervalle confiance est calculé par :

$$CI_i = t(0{,}05; M)\frac{\sigma}{\sqrt{M}} \tag{4.2}$$

où $t(0{,}05; M)$ correspond au 95^{e}percentile de la distribution de *Student*. Ainsi, CI_i quantifie la précision pour une condition donnée (mesure inter-sujet).

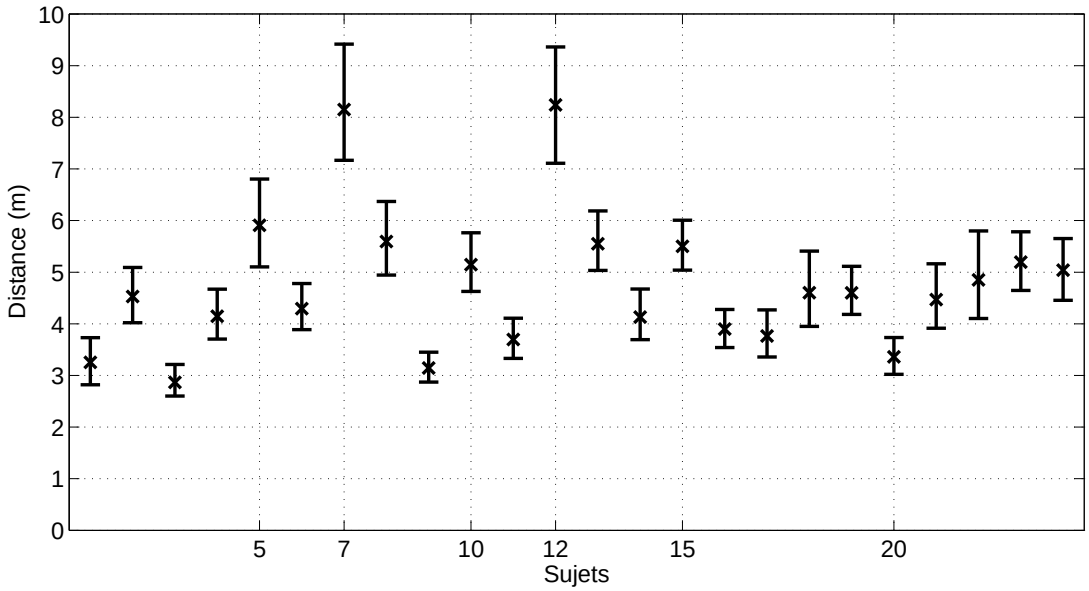


Figure 4.1 – Distance perçue moyenne pour chaque sujet. Les barres d'erreurs représente l'intervalle de confiance à 95%.

4.1.3 Analyse des jugements par sujet

Les résultats pour les 24 sujets sont analysés afin d'évaluer la cohérence de chaque sujet. Deux paramètres statistiques ont été calculés :

- la moyenne globale du sujet pour toutes les conditions avec indices auditifs et visuels cohérents (1 à 40),
- la moyenne de l'écart-type par condition.

La figure 4.1 montre que la distance absolue varie entre les sujets, entre 3 et 8 m. Cette variation entre les sujets a été observée dans d'autres études, voir par exemple Begault [1992]. Les deux sujets 7 et 12 estiment des distances plus grandes que les autres sujets et ils sont les moins cohérents dans leurs jugements, voir figure 4.2. Ces deux sujets ont donc été exclus de l'analyse des résultats.

4.1.4 Analyse de la distribution des jugements

Afin d'appliquer les mesures statistiques habituelles comme l'analyse de variance, la distribution des jugements doit suivre une loi dite normale ou Gaussienne. Cette loi définit la distribution y à tout point x d'une échelle :

$$y = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right) \tag{4.3}$$

où μ définit la moyenne de la distribution et σ définit l'écart-type de cette distribution.

La distribution des jugements des sujets a été analysée avec deux paramètres et un test statistique. Les deux paramètres définissant la distribution sont les coefficients de *Skewness* et de *Kurtosis*. Le premier paramètre, *Skewness*, définie l'asymétrie de la distribution

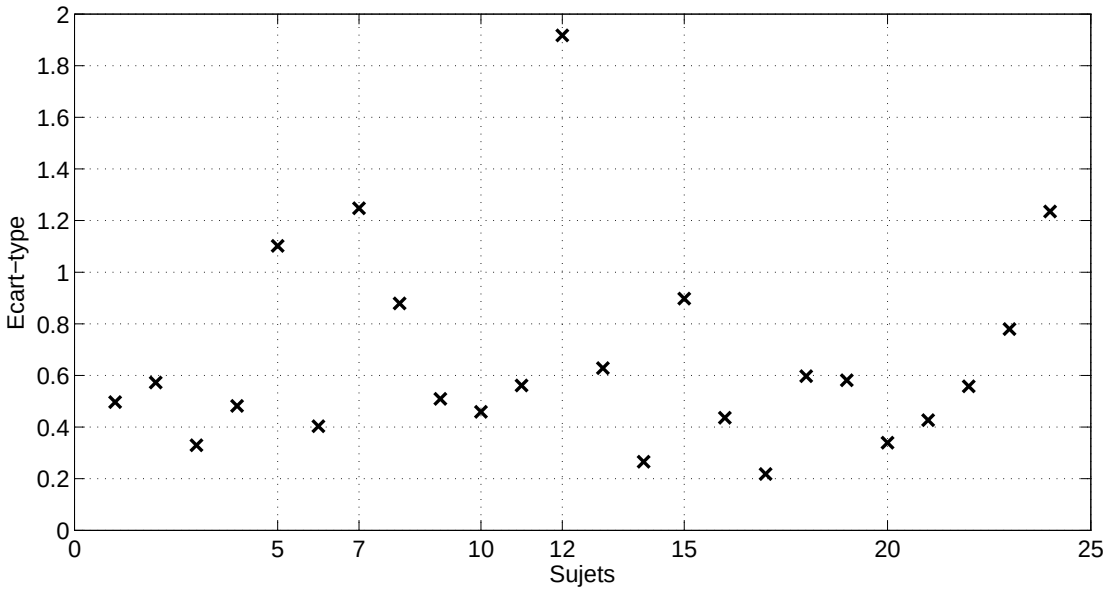


Figure 4.2 – Moyenne de l'écart-type pour chaque sujet.

autour de la moyenne, μ. Le second paramètre, *Kurtosis*, quantifie l'étalement de la distribution. Pour respecter la loi normale, la coefficient de *Skewness* doit être compris entre -2 et 2 et la coefficient de *Kurtosis* doit être inférieur à 3.

Ces deux coefficients ont été calculés pour chaque distance cible, 2, 3, 5, 10 et 20 m, en regroupant les jugements pour les conditions auditives, visuelles et bimodales. Le tableau 4.1 montre les valeurs des deux paramètres pour chaque distance cible. Les valeurs de *Skewness* sont supérieures à 0 ce qui indique une asymétrie vers les distances supérieures. Le même phénomène est accentué lors d'une analyse par condition. Ce phénomène est observé sur les histogrammes des jugements des sujets. La figure 4.3 montre cet histogramme pour la distance cible 3 m. De plus, les valeurs de *Kurtosis* ne suivent pas celles d'une distribution normale. En effet, de nombreux sujets ont donné une valeur entière pour la distance cible. Par exemple, 43,8% des sujets ont perçu la distance de l'objet à 2,0 m pour la distance cible 3 m.

Tableau 4.1 – Coefficients de *Skewness* et et de *Kurtosis* pour chaque distance cible.

Distance (m)	*Skewness*	*Kurtosis*
2	1,82	9,87
3	2,12	12,84
5	1,52	7,13
10	0,61	3,74
20	0,46	3,13

Par conséquent, nous pouvons conclure que la distribution des jugements des sujets ne suit pas la loi normale. Les analyses statistiques paramétriques comme l'Analyse de

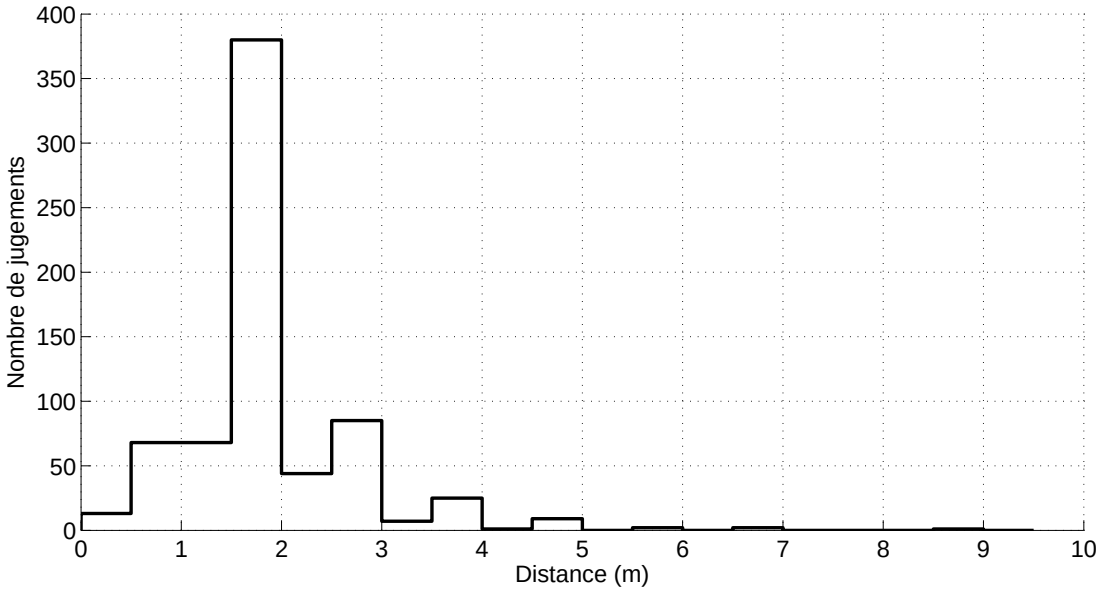


Figure 4.3 – Histogramme des jugements des sujets pour la distance cible 3 m.

Variance *analysis of variance* (ANOVA) ne sont donc pas appropriés et Il est nécessaire d'utiliser des mesures statistiques non-paramétriques :

- L'intervalle de confiance à 95%, *CI*, est calculé en utilisant la technique *Bootstrap* qui est une méthode d'inférence statistique (quelques jugements induisent les caractéristiques d'un ensemble plus vaste). Cette méthode se base sur un ré-échantillonnage des jugements (limitée à 2000 échantillons dans notre cas). L'intervalle de confiance est ensuite délimité par les percentiles 2,5% et 97,5% de la distribution obtenue. En particulier cette technique nous permet d'obtenir un intervalle de confiance asymétrique.
- L'ANOVA de *Kruskal-Wallis* est calculée pour évaluer l'effet des sujets (intergroupe), i.e. des échantillons indépendants.
- L'ANOVA de *Friedman* est calculée pour évaluer l'effet des conditions (intragroupe), i.e. des échantillons appariés.

4.2 Conditions avec indices cohérents

4.2.1 Effet de l'ordre des blocs

Nous pouvons supposer que le fait de visualiser l'environnement virtuel apporte des a priori sur l'acoustique de l'environnement et donne des indices pour la localisation en distance des conditions avec audition seule. En effet, [Richardson and Waller, 2005] ont montré qu'un simple entraînement des sujets de 5 à 7 minutes dans un environnement virtuel, ou un retour de l'examinateur sur leur performance, permettait de réduire la sous-estimation de la distance. De même, Plumert et al. [2005] ont montré qu'une première expérience de localisation en distance dans un environnement réel avait une influence sur une seconde expérience de localisation dans un EV reproduisant l'environnement réel. Les sujets n'ayant pas fait la première expérience avec environnement réel obtiennent de moins bon résultats

(distance perçue plus faible) que les sujets ayant eu une pré-visualisation de l'environnement réel.

Les sujets sont susceptibles d'estimer différemment la distance de l'objet s'ils ont jugé le bloc visuel ou le bloc auditif en premier. Cet effet est intergroupe (échantillons indépendants) car chaque groupe de sujets a évalué les blocs dans un ordre différent. C'est donc le test de *Kruskal-Wallis* qui a été appliqué sur les jugements obtenus pour la première session afin de mesurer l'impact de l'ordre des blocs. Le résultat du test montre que l'ordre des blocs, toutes conditions confondues, n'a aucun impact sur la distance perçue (χ^2 = 0,04 ; p = 0,84). Cependant, un test de *Kruskal-Wallis* par distance et par modalité montre un effet significatif de l'ordre des blocs pour seulement trois des vingt conditions :

15 : (environnement visuel pauvre, 20 m) χ^2 = 4,17 ; $p < 0{,}05$,

16 : (environnement visuel riche, 2 m) χ^2 = 4,56 ; $p < 0{,}05$,

20 : (environnement visuel riche, 20 m) χ^2 = 5,51 ; $p < 0{,}05$.

Pour les distances cibles 20 m, la distance perçue est significativement plus faible pour les sujets ayant passé le bloc vision en premier. On observe l'effet inverse pour la distance cible 2 m dans un environnement visuel riche, voir figure 4.4.

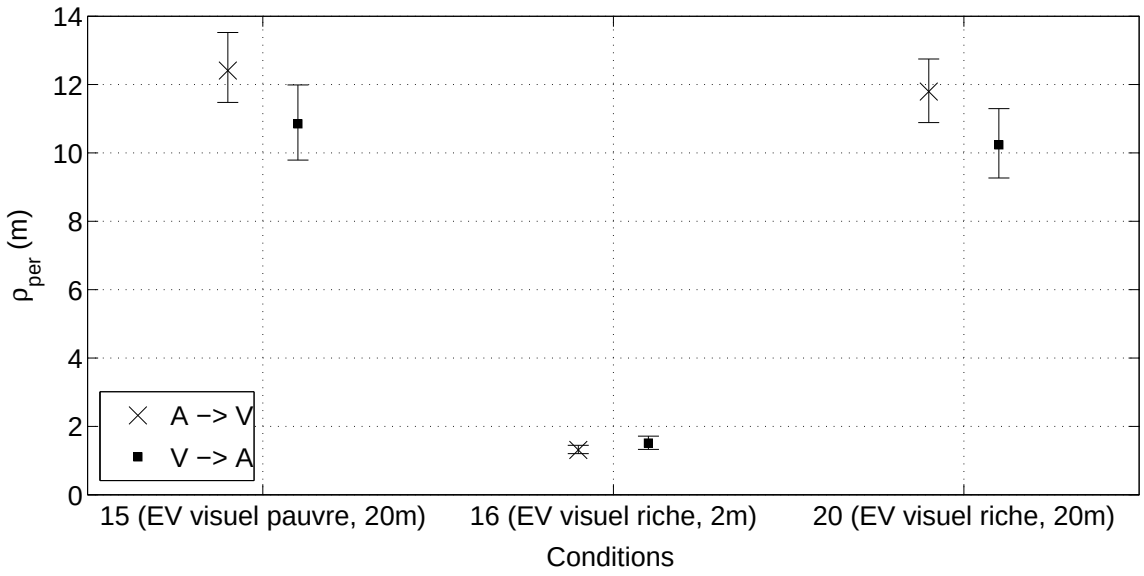


Figure 4.4 – Influence de l'ordre des blocs. A -> V ≡ modalité auditive en premier. V -> A ≡ modalité visuelle en premier.

4.2.2 Modalité auditive

La figure 4.5 montre la relation entre la distance cible, $\rho_{cib,i}$, et la distance perçue, $\rho_{per,i}$, par les sujets pour les conditions auditives ($i \in$ 1–10). Les barres d'erreurs représente l'intervalle de confiance à 95% calculé en utilisant la technique *Bootstrap*. Les courbes en gris représentent la fonction interpolée à partir des résultats, voir équation (1.1). Les résultats sont cohérents avec la littérature (voir section 1.2) : pour le temps de réverbération de T_{60} = 860 ms, les sujets estiment fidèlement les distances cibles 2, 3 et 5 m. Cependant, pour les deux autres distances cibles, et pour le temps de réverbération de T_{60} = 370 ms, les

distances cibles sont largement sous-estimées. De plus, la variabilité dans les jugements individuels (i.e. intervalle de confiance) est assez grande et augmente avec la distance cible. Des résultats similaires ont été observés par Zahorik et al. [2005].

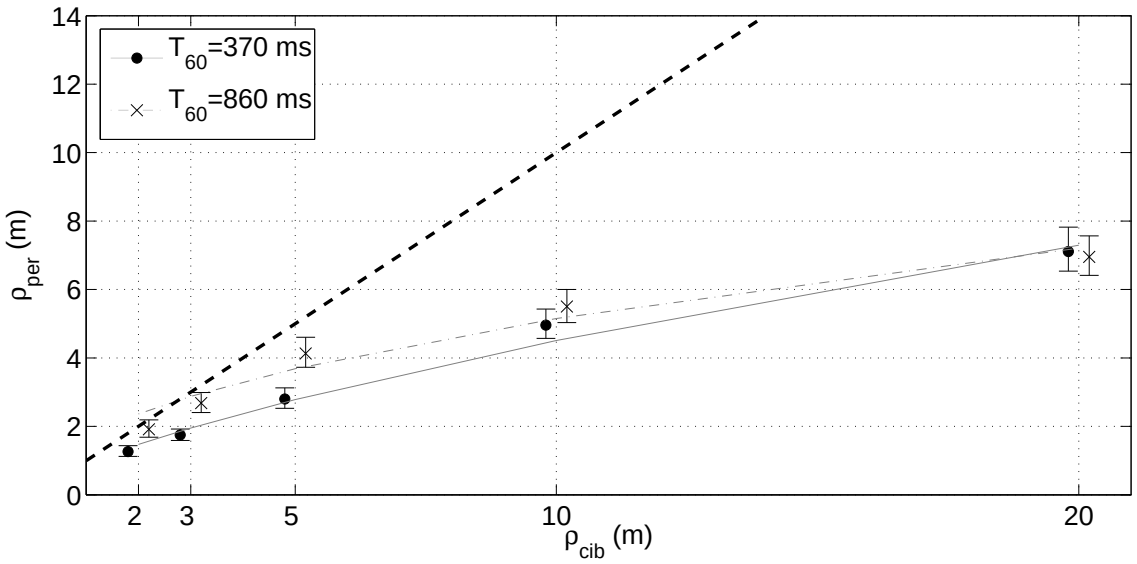


Figure 4.5 – Relation entre la distance auditive perçue, $\rho_{per,i}$, et la distance cible, $\rho_{cib,i}$.

Les jugements des sujets ont été analysés par un test de *Friedman* (i.e. échantillons appariés) avec les variables fixes (intra-sujet) *réverbération* et *distance cible* et 22 répétitions (sujets). Le test montre que la réverbération (χ^2 = 17,5 ; $p < 10^{-4}$) et la distance cible (χ^2 = 146,49 ; p = 0) ont une influence significative sur la distance perçue.

Un test de *Friedman*, par temps de réverbération, montre que la distance cible à une influence pour chaque réverbération (χ^2 = 84,8, p = 0 pour T_{60} = 370 ms, χ^2 = 79,5, p = 0 pour T_{60} = 860 ms). Cependant, la présentation de toutes les distances cibles dans un unique bloc donne aux sujets des informations de distances relatives entre les différentes conditions [Mershon and Bowers, 1979]. Cet effet a probablement facilité la différenciation des distances cibles.

Un test de *Friedman*, par distance cible, montre que la réverbération a une influence sur la distance perçue pour les distances cibles 2, 3 et 5 m uniquement. Ces résultats sont en accord avec l'hypothèse faîte par Bronkhorst and Houtgast [1999] sur l'influence du ratio entre énergie du champ direct et celle du champ diffus. En effet, une augmentation du temps de réverbération induit une augmentation de la distance perçue, voir figure 4.5. Ces résultats montrent que les deux indices acoustiques, (i) niveau du champ direct et (ii) ratio champ direct/champ diffus, ont une influence sur la distance égocentrique perçue par les sujets.

4.2.3 Modalité visuelle

La figure 4.6 montre la relation entre la distance cible, $\rho_{cib,i}$, et la distance perçue, $\rho_{per,i}$, par les sujets pour les conditions visuelles ($i \in$ 11–20). Les barres d'erreurs représentent

l'intervalle de confiance à 95% calculé en utilisant la technique *Bootstrap*. Les courbes en gris représentent la fonction interpolée à partir des résultats, voir équation (1.1). Les résultats sont cohérents avec la littérature (voir section 1.3) et la modalité auditive : les distances cibles sont largement sous-estimées pour les deux EV et pour toutes les distances cibles. De plus, comme pour la modalité auditive, la variabilité dans les jugements individuels augmente avec la distance cible. Cependant, on peut noter que cette variabilité est plus faible que pour la modalité auditive. Ainsi, la modalité visuelle permet une meilleure précision dans l'estimation de la distance que la modalité auditive.

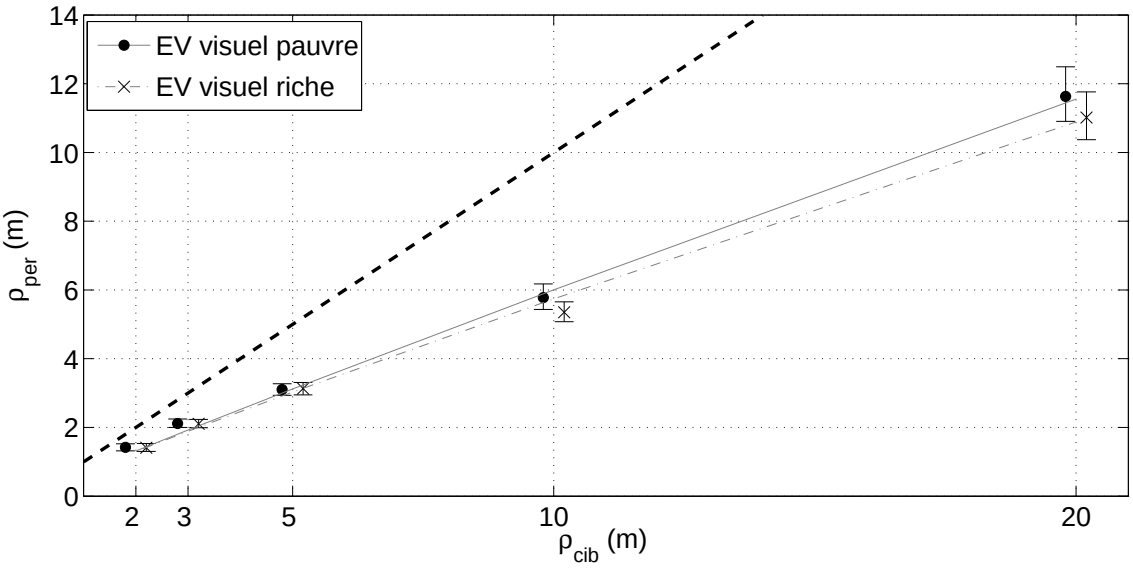


Figure 4.6 – Relation entre la distance visuelle perçue, $\rho_{per,i}$, et la distance cible, $\rho_{cib,i}$.

Les jugements des sujets ont été analysés par un test de *Friedman* (i.e. échantillons appariés) avec les variables fixes (intra-sujet) *environnement* et *distance cible* et 22 répétitions (sujets). Les résultats sont différents de la modalité auditive. En effet, toutes conditions confondues, la distance cible a une influence significative sur les jugements des sujets (χ^2 = 191,65 ; p = 0), contrairement à l'environnement visuel (χ^2 = 0,45 ; p = 0.50).

Un test de *Friedman*, par environnement, montre que la distance cible a une influence dans chaque environnement visuel : χ^2 = 87,2 ; p = 0 pour l'environnement pauvre en indices visuels et χ^2 = 88,0 ; p = 0 pour l'environnement riche en indices visuels.

Cependant, un test de *Friedman*, par distance cible, montre une influence de l'environnement visuel pour les distances cibles 10 et 20 m uniquement. Ce résultat montre que, contrairement à la modalité auditive, c'est l'objet visuel et non son environnement qui est la source principale de décision lorsque l'objet est proche du sujet. En effet, l'introduction d'indices visuels supplémentaires dans l'environnement ne permet pas de modifier la distance perçue bien que certaines différences soient visibles pour les grandes distances. Il est probable que les différences visuelles entre les deux environnements utilisés dans cette expérience ne soient pas assez nombreuses pour obtenir un impact sur la distance perçue.

Puisque pour les distances supérieures à 5–6 m, la convergence est peu efficace, les indices visuels utilisés par les sujets sont la perspective, la familiarité avec la source visuelle (haut-parleur) et la disparité binoculaire.

Il serait intéressant de reproduire cette expérience en utilisant la même source visuelle dans un environnement vide d'ancres visuelles (i.e. avec uniquement un horizon).

4.2.4 Bimodalité

La figure 4.7 montre la relation entre la distance cible, $\rho_{cib,i}$, et la distance perçue, $\rho_{per,i}$, par les sujets pour les conditions bimodales ($i \in$ 21–40). Les barres d'erreurs représentent l'intervalle de confiance à 95% calculé en utilisant la technique *Bootstrap*. Les courbes en gris représentent la fonction interpolée à partir des résultats, voir équation (1.1). Les résultats sont similaires à la modalité visuelle : les distances cibles sont largement sous-estimées pour les deux EV, pour les deux réverbérations et pour toutes les distances cibles. Comme pour la modalité visuelle, la variabilité dans les jugements individuels est plus faible que pour la modalité auditive.

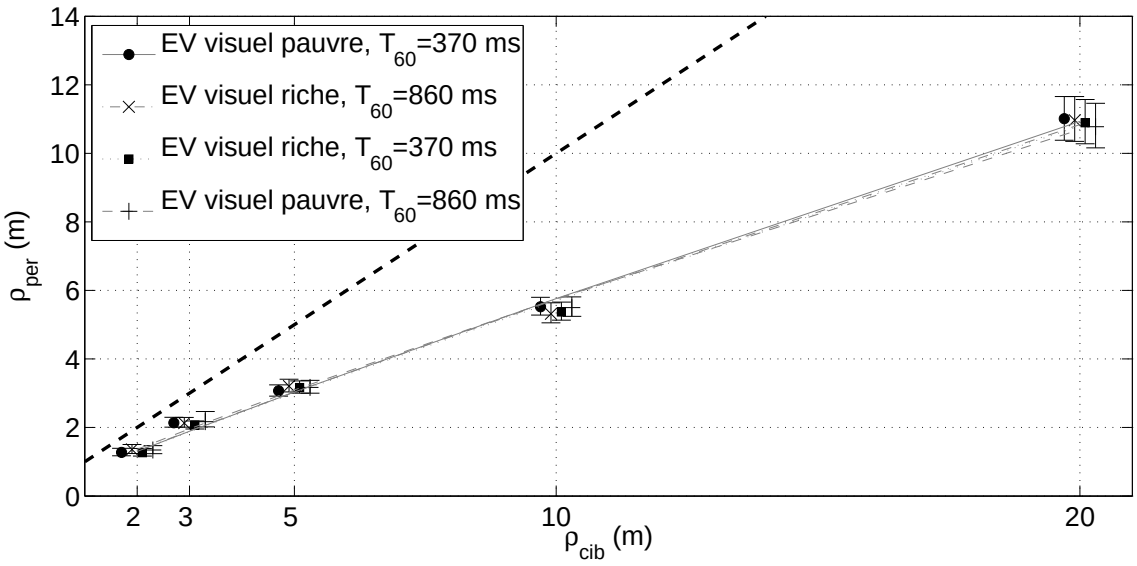


Figure 4.7 – Relation entre la distance perçue, $\rho_{per,i}$, et la distance cible, $\rho_{cib,i}$, pour les conditions bimodales.

Les jugements des sujets ont été analysés par un test de *Friedman* (i.e. échantillons appariés) avec les variables fixes (intra-sujet) *environnement* (i.e. deux environnements visuels * deux temps de réverbération) et *distance cible* et 22 répétitions (sujets). Bien que la distance cible ait une influence significative sur les jugements des sujets (χ^2 = 396,08 ; p = 0), l'environnement (χ^2 = 0,25 ; p = 0,97) n'a pas d'influence sur la distance perçue. Cependant, un test de *Friedman*, avec les variables environnement, sujets et 4 répétitions, montre une influence significative de l'environnement pour la distance cible 10 m (χ^2 = 10,5 ; $p < 0{,}05$).

Un test de *Friedman* montre que pour les quatre environnements, la distance cible a une influence sur la distance perçue :

- $\chi^2 = 87{,}2$; $p = 0$ pour $T_{60} = 370$ ms et environnement visuel pauvre,
- $\chi^2 = 88{,}0$; $p = 0$ pour $T_{60} = 860$ ms et environnement visuel riche,
- $\chi^2 = 88{,}0$; $p = 0$ pour $T_{60} = 370$ ms et environnement visuel riche,
- $\chi^2 = 88{,}0$; $p = 0$ pour $T_{60} = 860$ ms et environnement visuel pauvre.

Il est difficile de conclure à partir de ces résultats, cependant, la réverbération semble avoir une influence réduite dans un environnement bimodal par rapport à un environnement purement auditif sonore. Ce résultat montre que l'introduction de la vision réduit l'impact de l'audition. Cet effet peut être considéré comme un effet d'*attraction* de la perception vers l'objet visuel.

4.2.5 Comparaison entre les différentes modalités

Vision *vs*. audition

Un test de *Friedman* a été utilisé avec les variables fixes (intra-sujet) *modalité* et *distance cible* * *environnements*, et avec 22 répétitions (sujets). Ce test montre une influence de la modalité dans le jugement des sujets ($\chi^2 = 9{,}14$; $p = 0{,}003$), toutes distances cibles et les deux environnements confondus. De plus, un test de *Friedman*, par distance cible, montre que la modalité a une influence sur la distance perçue pour la distance cible la plus lointaine (20 m) $\chi^2 = 168{,}22$; $p = 0$ uniquement. En effet, pour la distance cible 20 m, la distance perçue est inférieure dans la modalité auditive seule par rapport à la modalité visuelle seule.

Audition *vs*. bimodal

Afin de tester l'influence de la modalité entre les conditions avec indices auditifs seuls et les conditions bimodales, nous utilisons deux tests de *Friedman*, un par environnement visuel. Le premier test est appliqué à l'environnement visuel pauvre et le deuxième à l'environnement visuel riche. Ces tests ont comme variable fixe (intra-sujet) *la modalité* et *distance cible* * *réverbération* et 22 répétitions (sujets). Les deux tests montrent, toutes distances cibles et les deux temps de réverbération confondus, une influence significative de la modalité :

- $\chi^2 = 8{,}22$, $p = 0{,}004$ pour l'environnement visuel pauvre,
- $\chi^2 = 6{,}17$, $p = 0{,}013$ pour l'environnement visuel riche.

Une analyse détaillée pour chaque condition montre que la modalité a une influence significative pour la distance cible 20 m (dans les deux environnements visuels).

Vision *vs*. bimodal

Les sections 4.2.3 et 4.2.4 ont montré des résultats similaires pour la modalité visuelle seule et les conditions avec indices visuels et auditifs. Comme pour l'audition, deux tests de *Friedman* sont utilisés : le premier test est appliqué au temps de réverbération de $T_{60} = 370$ ms et le deuxième au temps de réverbération de $T_{60} = 860$ ms. Ces deux tests ont

comme variables la *modalité*, la *distance cible * environnement visuel* et 22 répétitions (sujets). Les deux tests montrent, toutes distances cibles et les deux environnements visuels confondus, que la modalité n'a aucune influence sur les jugements des sujets :

- $\chi^2 = 1{,}74$, $p = 0{,}19$ pour le temps de réverbération $T_{60} = 370$ ms,
- $\chi^2 = 0{,}6$, $p = 0{,}44$ pour le temps de réverbération $T_{60} = 860$ ms.

En effet, une analyse détaillée par condition, montre que la modalité n'a aucune influence quelque soit la distance cible.

4.3 Conditions avec indices incohérents

Le corpus de test présenté dans la section 3.1.1 inclut huit conditions bimodales avec indices visuels et sonores incohérents, voir tableau 3.4. Ces huit conditions sont donc définies par la distance cible de la source visuelle, $\rho_{cib,V}$, et la distance cible de la source sonore, $\rho_{cib,A}$. Plusieurs tests statistiques sont utilisés afin de déterminer si l'écart entre la position de la source sonore et celle de la source visuelle a un impact sur la distance perçue :

1. Nous vérifions d'abord si l'écart, introduit entre les indices visuels et sonores pour les conditions avec indices incohérents, peut avoir une influence sur les conditions avec indices cohérents. Ainsi nous comparons (i) les conditions avec indices cohérents correspondant à la position de la source visuelle, $\rho_{cib,V}$, (ii) aux conditions avec indices cohérents correspondant à la position de la source sonore, $\rho_{cib,A}$. Les résultats sont montrés dans le tableau 4.2.
2. Nous comparons ensuite (i) les conditions avec indices incohérents (ii) aux conditions avec indices cohérents qui correspondent à la position de la source sonore, $\rho_{cib,A}$.
3. Enfin, nous comparons (i) les conditions avec indices incohérents (ii) aux conditions avec indices cohérents qui correspondent à la position de la source visuelle, $\rho_{cib,V}$. Les résultats sont montrés dans le tableau 4.3.

Dans les tableaux 4.2 et 4.3, la distance (2e colonne) correspond à la position de la source visuelle. L'offset (3e colonne) détermine la distance entre la position de la source visuelle et la position de la source sonore.

Nous comparons dans un premier temps les conditions avec indices cohérents correspondant à la position de la source visuelle, $\rho_{cib,V}$, aux conditions avec indices cohérents correspondant à la position de la source sonore, $\rho_{cib,A}$. Cependant, les sujets n'ont pas évalué les conditions avec indices cohérents correspondant à la position de la source visuelle, $\rho_{cib,V}$, des conditions avec indices incohérents 41, 43 et 47. En effet, les conditions avec indices cohérents n'ont pas été évalué aux distances cibles 1 et 15 m. Les tests de *Friedman* montrent qu'une modification de la distance cible des conditions avec indices cohérents correspondant à la distance entre $\rho_{cib,V}$ et $\rho_{cib,A}$ a une influence sur la distance perçue, voir tableau 4.2.

Nous comparons ensuite les conditions avec indices incohérents aux conditions avec indices cohérents correspondant à la position de la source sonore, $\rho_{cib,A}$. Cependant, il n'est pas possible de comparer les conditions 41, 43 et 47 aux conditions avec indices cohérents car ces dernières n'ont pas été évaluées aux distances cibles 1 et 15 m. La distance perçue

Tableau 4.2 – Résultats des tests de *Friedman* pour la comparaison des conditions avec indices cohérents ($\rho_{cib,V}$ *vs.* $\rho_{cib,V}$). Le symbole ***** montre les conditions significativement différentes.

Conditions	Distance (m)	Offset (m)	p
42 : 33 *vs.* 35	5	+15	0*****
44 : 33 *vs.* 34	5	+5	0*****
45 : 34 *vs.* 35	10	+10	0*****
46 : 34 *vs.* 32	10	-7	0*****
48 : 35 *vs.* 33[a]	20	-15	0*****

[a] Comparaison équivalente à la première ligne.

pour chaque condition 42, 44, 45, 46 et 48 est significativement différente de la distance perçue pour la condition correspondant à la position de la source sonore (i.e. condition 35, 34, 35, 32 et 33, respectivement). Ainsi, lorsque la source visuelle n'est pas positionnée au même endroit que la source sonore, les sujets ont tendance à percevoir l'objet (multimodal) au niveau de la source visuelle. Ce résultat correspond à l'effet "ventriloque" introduit dans la section 1.4. Lorsque les indices visuels et auditifs ne sont pas cohérents (dans l'espace), la source sonore est déplacée vers la source visuelle. Par exemple Lewald et al. [2001] a étudié l'effet ventriloque en azimut. Cependant, même si Howard and Templeton [1966] a montré que cet effet existait en distance, peu d'études ont montré ce phénomène.

Enfin, nous comparons les conditions avec indices incohérents aux conditions avec indices cohérents correspondant à la position de la source visuelle, $\rho_{cib,V}$. La figure 4.8 montre l'influence de la position de la source sonore pour une source visuelle positionnée à 5 m. On remarque que les seules conditions avec indices incohérents qui sont significativement différentes des conditions avec indices cohérents sont celles avec la source sonore placée devant la source visuelle (i.e. conditions 43, 46 et 48), sauf pour la distance cible la plus proche (2 m), voir tableau 4.3. Ces résultats semblent montrer que l'effet ventriloque est asymétrique et n'apparaît que pour des sources sonores placées derrière la source visuelle. Cependant, cette asymétrie est influencée par la distance de la source visuelle. En effet, pour une source visuelle à 2 m du sujet, un offset de -1 m ne permet pas de supprimer cet effet ventriloque. De plus, un offset de -4 m permet de supprimer l'effet ventriloque pour une source visuelle à 5 m mais vraisemblablement pas pour une source visuelle à 20 m. Il sera donc nécessaire de quantifier avec précision l'impact de cet effet ventriloque pour différentes distances cibles et différents offsets.

4.4 Modélisation

A partir des distances perçues par les sujets et des distances cibles, il est possible de modéliser la perception en distance d'un objet, i.e. $\rho_{per,i} = f(\rho_{cib,i})$. Da Silva [1985] a montré que la distance perçue pouvait être modélisée par la loi de *Stevens* décrite en section 1.1. Pour chaque modalité, les coefficients k et a de l'équation (1.1) sont estimés au sens des moindres carrés. Le coefficient a représente le taux de compression de la distance cible tandis que le coefficient k définit la sur- ou sous-estimation de la distance cible. A partir de

Tableau 4.3 – Résultats des tests de *Friedman* pour la comparaison des conditions avec indices incohérents (41 à 48) aux conditions avec indices cohérents correspondant à la position de la source visuelle. Le symbole ***** montre les conditions significativement différentes.

Conditions	Distance (m)	Offset (m)	*p*
41 *vs.* 31	2	-1	,564
42 *vs.* 33	5	+15	,999
43 *vs.* 33	5	-4	,001*****
44 *vs.* 33	5	+5	,796
45 *vs.* 34	10	+10	,467
46 *vs.* 34	10	-7	,002*****
47 *vs.* 34	10	+5	,492
48 *vs.* 35	20	-15	,002*****

la loi de *Stevens*, des valeurs sont estimées, $\rho_{est,i}$, puis comparées aux distances perçues, $\rho_{per,i}$. La racine de l'erreur quadratique moyenne est calculée pour chaque modalité à partir des valeurs estimées et des distances perçues par les sujets. Les résultats sont présentés dans le tableau 4.4. Les valeurs entre parenthèses correspondent aux intervalles de confiance à 95% pour les coefficients *k* et *a*.

Tableau 4.4 – Coefficients utilisés pour la modélisation de la perception en distance pour chaque modalité. *RMSE* ≡ racine de l'erreur quadratique moyenne. EV ≡ environnement visuel.

Modalité	Description	*k*	*a*	*RMSE*
Auditive	T_{60} = 370 ms	0,91 (±0,43)	0,69 (±0,18)	0,29
Auditive	T_{60} = 860 ms	1,69 (±0,81)	0,48 (±0,19)	0,40
Visuelle	EV pauvre	0,68 (±0,17)	0,92 (±0,09)	0,16
Visuelle	EV riche	0,68 (±0,26)	0,92 (±0,14)	0,24
Bimodale	T_{60} = 370 ms, EV pauvre	0,69 (±0,19)	0,92 (±0,10)	0,18
Bimodale	T_{60} = 860 ms, EV riche	0,69 (±0,28)	0,92 (±0,15)	0,26
Bimodale	T_{60} = 370 ms, EV riche	0,68 (±0,22)	0,92 (±0,12)	0,21
Bimodale	T_{60} = 860 ms, EV pauvre	0,73 (±0,19)	0,89 (±0,10)	0,17

On remarque que pour la modalité auditive, les coefficients *a* sont inférieurs à 1 ce qui indique une forte compression de la distance cible. En effet, les courbes grises sur la figure 4.5 montrent une augmentation de la sous-estimation lorsque la distance cible augmente. Ce résultat est cohérent avec la littérature. Par exemple, Zahorik [2001] obtient un coefficient de compression de $a = 0{,}66$ pour une expérience utilisant des haut-parleurs réels et non virtuels. Cette compression est essentiellement due aux distances cibles 10 et 20 m. De plus, la variabilité dans les jugements des sujets pour cette modalité entraîne une erreur (*RMSE*) plus élevée que pour les autres modalités.

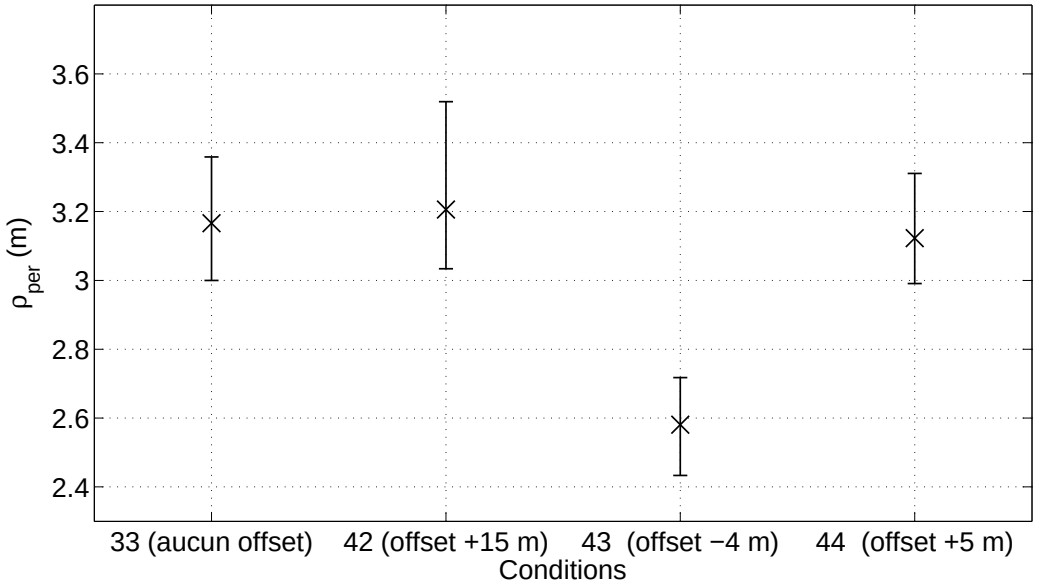


Figure 4.8 – Distance perçue, $\rho_{per,i}$, pour une source visuelle positionnée à 5 m du sujet et une source sonore cohérente (condition 33) ou une source sonore incohérente (conditions 42, 43 et 44).

Exceptées les conditions avec réverbération de T_{60} = 860 ms et un environnement pauvre en indices visuels, les autres conditions visuelles et bimodales produisent les mêmes coefficients. Les coefficients de compression, *a*, sont plus proches de 1 pour la modalité visuelle et les conditions bimodales. Cela indique que la distance cible n'est pas compressée mais que la relation entre distance perçue et distance cible est linéaire. Cependant, les coefficients *k* sont inférieurs à 1 ce qui indique une sous-estimation de la distance cible.

Les valeurs de *RMSE* pour l'environnement riche en indices visuels sont supérieures aux valeurs de *RMSE* pour l'environnement pauvre en indices visuels. La même tendance est observée pour les intervalles de confiance des coefficients *k* et *a*. Cette effet montre une plus grande variabilité des jugements des sujets pour l'environnement riche en indices visuels.

CONCLUSIONS

Cette étude a utilisé des systèmes de réalité virtuelle pour l'étude de la perception de la distance égocentrique. Cette étude se focalisait sur la comparaison de conditions unimodales (visuelles ou auditives) et bimodales (visuelles et auditives). Les résultats montrent que l'utilisation d'indices visuels ou auditifs pertinents permet de varier la perception en profondeur dans des applications de réalité virtuelle. En effet, pour chaque environnement, les sujets ont réussi à différencier les différentes distances cibles. Cependant, la modalité visuelle induit une meilleure précision de la localisation que la modalité auditive.

Les résultats sont cohérents avec les performances observées dans un environnement réel : les distance cibles sont sous-estimées pour quasiment toutes les conditions avec un facteur proche de 2. De plus, on observe une augmentation de la variabilité intra-sujet lorsque la distance cible augmente.

Les résultats sont différents suivant la modalité. En effet, la relation entre la distance perçue par les sujets et la distance cible est différente suivant la modalité. En effet, cette relation est linéaire dans les cas modalité visuelle seule et bimodalité et courbe dans le cas modalité auditive seule. La modalité auditive seule induit une plus grande sous-estimation et une moins bonne précision de la localisation que la modalité visuelle seule. Cependant, les résultats pour la modalité visuelle seule sont assez proches des résultats pour les conditions bimodales. Cet effet suggère que la combinaison de la modalité visuelle et auditive n'induit pas une meilleure estimation de la distance que la modalité visuelle seule.

Enfin, les résultats pour les conditions avec indices visuels et auditifs incohérents ont montré un effet ventriloque en profondeur. Cependant, cet effet est visible lorsque l'indice auditif est placé derrière l'indice visuel ce qui implique une effet ventriloque asymétrique. cependant, il serait intéressant de quantifier avec plus de précision cet effet ventriloque.

Enfin, cette étude est basée sur une estimation directe de la distance égocentrique pour des objets statiques. D'autres méthodes de mesure, comme une tâche de triangularisation devraient être utilisées pour confirmer les résultats de cette étude.

TEXTE INTRODUCTIF DESTINÉ AUX SUJETS

Distance d'un objet virtuel sonore et/ou visuel

Bonjour,
Tout d'abord, nous vous remercions d'avoir accepté de participer à notre test de réalité virtuelle. Veuillez prendre quelques minutes pour lire ces instructions. N'hésitez pas à poser vos questions à l'expérimentateur à la fin de votre lecture.
Vous participez aujourd'hui à un test perceptif où vous devez estimer la distance qui vous sépare d'un objet virtuel.
Après la lecture de ce texte, vous aurez à observer et/ou écouter différentes scènes virtuelles. Elles correspondent à de courtes phrases diffusées par un haut-parleur.
Déroulement du test :
Vous allez d'abord, durant un courte phase d'entraînement, observer quelques exemples de scènes virtuelles. Viendront ensuite les séances d'une durée de 10 à 15 minutes. De courtes pauses vous serons proposées entre chaque séance.
A la fin de chaque séquence sonore/visuelle, il vous sera demandé d'estimer la distance de l'objet virtuel :

Quelle est la distance de l'objet virtuel :

Cette distance doit être positive ou nulle, et peut avoir une précision d'un chiffre après la virgule. Exemples : 5,8 ou 5
La séquence suivante est jouée automatiquement après avoir appuyé sur la touche "Enter" du pavé numérique.
A la fin du test, il vous sera demandé de contacter l'expérimentateur en utilisant le téléphone situé dans le hall du bâtiment. Le numéro est le 89 71.
Veuillez prendre en compte les critères suivants lors de votre jugement :

Il est important de vous imaginer dans une scène réelle !

Vous devez observer l'objet virtuel comme étant un objet réel positionné dans une pièce réelle. Vous devez donc estimer la distance de manière **intuitive**.
De plus cette étude est de nature purement subjective, **il n'y a donc pas de réponse correcte ou fausse**. Seule votre impression personnelle sur la distance est importante.
Si vous deviez avoir des difficultés durant le test perceptif, veuillez prévenir l'expérimentateur. Celui-ci est à votre disposition pour répondre à vos questions.

Bibliographie

J. Andre and S. Rogers. Using Verbal and Blind-Walking Distance Estimates to Investigate the Two Visual Systems Hypothesis. *Attention, Perception, & Psychophysics*, 68(3) : 353–361, 2006. 3, 12

D. Ashmead, D. Davis, and A. Northington. Contribution of Listeners' Approaching Motion to Auditory Distance Perception. *Journal of Experimental Psychology : Human Perception and Performance*, 21(2) :239–256, 1995. 3, 8

D.R. Begault. Preferred Sound Intensity Increase for Sensation of Half Distance. *Perceptual Motor Skills*, 72(3) :1019–1029, 1991. 5

D.R. Begault. Perceptual Effects of Synthetic Reverberation on Three-Dimensional Audio Systems. *The Journal of the Audio Engineering Society*, 40 :895–904, 1992. 25, 27, 42

D.R. Begault, E.M. Wenzel, and M.R. Anderson. Direct Comparison of the Impact of Head Tracking, Reverberation, and Individualized Head-Related Transfer Functions on the Spatial Perception of a Virtual Speech Source. *The Journal of the Audio Engineering Society*, 49(10) :904–916, 2001. 25, 28

G. Békésy and E.G. Wever. *Experiments in Hearing*. McGraw-Hill, Oxford, England, 1960. 2

A.J. Berkhout, D. de Vries, and P. Vogel. Acoustic Control by Wave Field Synthesis. *The Journal of the Acoustical Society of America*, 93(5) :2764–2778, 1993. 26

J. Blauert. *Spatial Hearing : the Psychophysics of Human Sound Localization*. MIT Press, USA–Cambridge, Mass., révisée edition, 1997. 6, 23

A.W. Bronkhorst. Localization of Real and Virtual Sound Sources. *The Journal of the Acoustical Society of America*, 98(5) :2542–2553, Nov. 1995. 22, 24, 35

A.W. Bronkhorst and T. Houtgast. Auditory Distance Perception in Rooms. *Nature*, 397 : 517–520, February 1999. Issue 6719. 5, 14, 46

R.A. Butler, E.T. Levy, and W.D. Neff. Apparent Distance of Sounds Recorded in Echoic and Anechoic Chambers. *Journal of Experimental Psychology : Human Perception and Performance*, 6(4) :745–750, 1980. 5

D.R. Campbell, K.J. Palomaki, and G.J. Brown. Roomsim, a Matlab Simulation of Shoebox Room Acoustics for use in Teaching and Research. *Computing and Information Systems*, 9(3) :48–51, 2005. 28, 36

CERV. *Atelier de Réalité Virtuelle*, 31 March 2011. URL . 34

P. Cochran, J. Throop, and WE Simpson. Estimation of Distance of a Source of Sound. *The American Journal of Psychology*, 81(2) :198–206, 1968. 5

P.D. Coleman. An Analysis of Cues to Auditory Depth Perception in Free Space. *Psychological Bulletin*, 60(3) :302–315, 1963. 5, 6

P.D. Coleman. Dual Rôle of Frequency Spectrum in Determination of Auditory Distance. *The Journal of the Acoustical Society of America*, 44(2) :631–632, 1968. 6, 7

N. Côté, V. Koehl, M. Paquier, and F. Devillers. Interaction between Auditory and Visual Distance Cues in Virtual Reality Applications. In *Proc of Forum Acusticum*, Aalborg, Danemark, 2011. vi

S.H. Creem-Regehr, P. Willemsen, A.A. Gooch, and W.B. Thompson. The Influence of Restricted Viewing Conditions on Egocentric Distance Perception : Implications for Real and Virtual Environments. *Perception*, 34(2) :191–204, 2005. 11, 19

J.E. Cutting and P.M. Vishton. *Perception of Space and Motion*, chapter Perceiving Layout and Knowing Distances : The Integration, Relative Potency, and Contextual Use of Different Information about Depth, pages 69–117. Academic Press, New-York, USA, 1995. 2, 9, 11

J.A. Da Silva. Scales for Perceived Egocentric Distance in a Large Open Field : Comparison of Three Psychophysical Methods. *The American Journal of Psychology*, 98(1) :119–144, 1985. 2, 3, 51

J. Daniel, R. Nicol, and S. Moreau. Further Investigations of High Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging. In *Proc. of the 114th AES Convention*, number 5788, NL–Amsterdam, 2003. Audio Engineering Society ; 1999. 26

L. Deng and D. O'Shaughnessy. *Speech Processing : a Dynamic and Optimization-Oriented Approach*. Marcel Dekker, Inc., USA–New-York, NY, 2003. 35

R.O. Duda and W.L. Martens. Range Dependence of the Response of a Spherical Head Model. *The Journal of the Acoustical Society of America*, 104(5) :3048–3058, 1998. 7

NI Durlach, A. Rigopulos, XD Pang, WS Woods, A. Kulkarni, HS Colburn, and EM Wenzel. On the Externalization of Auditory Images. *Presence : Teleoperators & Virtual Environments*, 1(2) :251–257, 1992. 27

S.R. Ellis, K. Mania, B.D. Adelstein, and M.I. Hill. Generalizeability of Latency Detection in a Variety of Virtual Environments. In *Proc. of the Human Factors and Ergonomics Society Annual Meeting*, volume 48, pages 2632–2636. Human Factors and Ergonomics Society, 2004. 29

P. Fuchs. Interfaces Visuelles. *Techniques de l'Ingénieur*, (TE5906), 2003. ISSN 1632-3823. 18, 21

W. Fujisaki, S. Shimojo, M. Kashino, and S. Nishida. Recalibration of Audiovisual Simultaneity. *Nature Neuroscience*, 7(7) :773–778, 2004. 14

M.B. Gardner. Proximity Image Effect in Sound Localization. *The Journal of the Acoustical Society of America*, 43 :163, 1968. 15

M.B. Gardner. Distance Estimation of 0˚or Apparent 0˚-Oriented Speech Signals in Anechoic Space. *The Journal of the Acoustical Society of America*, 45(1) :47–53, 1969. 5, 35

W.C. Gogel and J.D. Tietz. Absolute Motion Parallax and the Specific Distance Tendency. *Attention, Perception, & Psychophysics*, 13(2) :284–292, 1973. 2, 3

G.R.A.S.©. *KEMAR Manikin Type 45BM and Type 45BA*, Retrieved 10 June 2011. URL . 36

R.L. Gregory. *Eye and brain : The Psychology of Seeing*. Princeton University Press, 5eedition, 1997. 11

J. Heron, D. Whitaker, P.V. McGraw, and K.V. Horoshenkov. Adaptation Minimizes Distance-related Audiovisual Delays. *The Journal of Vision*, 7(13) :1–8, 2007. ISSN 1534-7362. 15

I.P. Howard and W.B. Templeton. *Human Spatial Orientation*. John Wiley and Sons, USA–New York, 1966. 15, 51

ITU–T Handbook on Telephonometry. International Telecommunication Union, CH–Geneva, 1992. 37

V. Interrante, B. Ries, and Anderson L. Distance Perception in Immersive Virtual Environments, Revisited. In *IEEE Virtual Reality Conference (VR 2006)*, pages 3–10, 25–29 March 2006. 21

ITU–R Rec. BS.775–2. Multichannel Stereophonic Sound System with and without Accompanying Picture, July 2006. 23

M. Jeub, M. Schafer, and P. Vary. A Binaural Room Impulse Response Database for the Evaluation of Dereverberation Algorithms. In *16th Int. Conf. on Digital Signal Proc.*, pages 1–5, 2009. 27, 37

O. Kirkeby, P. A. Nelson, and H. Hamada. The “Stereo Dipole” : Binaural Sound Reproduction Using Two Closely Spaced Loudspeakers. In *Audio Engineering Society Convention 102*, 3 1997. 25

E. Klein, J.E. Swan, G.S. Schmidt, M.A. Livingston, and O.G. Staadt. Measurement Protocols for Medium-Field Distance Perception in Large-Screen Immersive Displays. In *IEEE Virtual Reality Conference, VR’2009*, pages 107–113, Lousiana, USA, March 14–18 2009. 3, 19

K. Koffka. *Principles of Gestalt Psychology*. Harcourt, Brace, 1935. 13

N. Kopčo, M. Schoolmaster, and B. Shinn-Cunningham. Learning to Judge Distance of Nearby Sounds in Reverberant and Anechoic Environments. In *Proceedings of the Joint Congress CFA/DAGA ’04*, Strasbourg, France, 22-25 March 2004. 6, 14

N. Kopčo, S. Santarelli, V. Best, and B. Shinn-Cunningham. Simulating distance cues in virtual reverberant environments. In *19th INTERNATIONAL CONGRESS ON ACOUSTICS*, MADRID, 2-7 SEPTEMBER 2007. 28

M. Lambooij, W. IJsselsteijn, M. Fortuin, and I. Heynderickx. Visual Discomfort and Visual Fatigue of Stereoscopic Displays : A Review. *The Journal of Imaging Science and Technology*, 53(3) :1–14, 2009. 22

M.S. Landy, L.T. Maloney, E.B. Johnston, and M. Young. Measurement and Modeling of Depth Cue Combination : In Defense of Weak Fusion. *Vision Research*, 35(3) :389–412, 1995. 12

J. Lewald, W.H. Ehrenstein, and R. Guski. Spatio-Temporal Constraints for Auditory-Visual Integration. *Behavioural Brain Research*, 121(1–2) :69–79, 2001. 14, 15, 51

T. Lokki, M. Grohn, L. Savioja, and T. Takala. A Case Study of Auditory Navigation in Virtual Acoustic Environments. In *Proc. of the Int. Conf. on Auditory Display (ICAD)*, pages 145–150, 2000. 27, 28, 35

J M Loomis, R L Klatzky, J W Philbeck, and R Gvariabil Golledge. Assessing Auditory Distance Perception Using Perceptually Directed Action. *Perception & Psychophysics*, 60(6) :966–980, 1998. 1, 3, 37

J.M. Loomis and J.M. Knapp. *Virtual and Adaptive Environments : Applications, Implications, and Human Performance Issues*, chapter Visual Perception of Egocentric Distance in Real and Virtual Environments, pages 21–46. Laurence Erlbaum Associates, NJ, USA, 2003. 17

P. Majdak, B. Laback, M. Goupell, and M. Mihocic. The Accuracy of Localizing Virtual Sound Sources : Effects of Pointing Method and Visual Environment. In *Proc. of the 124th AES Convention*, NL–Amsterdam, May 17–20 2008. 8, 14

M.W. Matlin and H.J. Foley. *Sensation and Perception*. Allyn and Bacon Boston, MA, Needhman Heights, Mass., 4^{e}edition, 1997. 13

B. Ménélas, L. Picinali, B.F.G. Katz, P. Bourdot, and M. Ammi. Haptic Audio Guidance for Target Selection in a Virtual Environment. In *Proc. of the 4th Int. Haptic and Auditory Interaction Design Workshop (HAID'09)*, volume 2, 2009. 1

D.H. Mershon and J.N. Bowers. Absolute and Relative Cues for the Auditory Perception of Egocentric Distance. *Perception*, 8(3) :311–322, 1979. 5, 46

D.H. Mershon and W.E. Hutson. Toward the Indirect Measurement of Perceived Auditory Distance. *Bulletin of the Psychonomic Society*, 29(2) :109–112, 1991. 3

D.H. Mershon and L.E. King. Intensity and Reverberation as Factors in the Auditory Perception of Egocentric Distance. *Attention, Perception & Psychophysics*, 18(6) :409–415, 1975. 4, 5

A. Murgia and P.M. Sharkey. Estimation of Distances in Virtual Environments Using Size Constancy. *The International Journal of Virtual Reality*, 1(8) :67–74, 2009. 17, 21

K.V. Nguyen, C. Suied, I. Viaud-Delmon, and O. Warusfel. Spatial Audition in a Static Virtual Environment : The Role of Auditory-Visual Interaction. *The Journal of Virtual Reality and Broadcasting*, 6(5), Mars 2009. 14

S.H. Nielsen. Auditory Distance Perception in Different Rooms. *The Journal of the Audio Engineering Society*, 41(10) :755–770, 1993. 5

M. Noisternig, T. Musil, A. Sontacchi, and R. Holdrich. 3D Binaural Sound Reproduction Using a Virtual Ambisonic Approach. In *Proc. of the Int. Symp. on Virtual Environments, Human-Computer Interfaces and Measurement Systems (VECIMS'03)*, pages 174–178, 2003. 27, 28

S. Palmisano, B. Gillam, D.G. Govan, R.S. Allison, and J.M. Harris. Stereoscopic Perception of Real Depths at Large Distances. *The Journal of Vision*, 10(6) :1–16, 2010. 11

R. Patterson. Human Factors of 3-D Displays. *The Journal of the Society for Information Display*, 15(11) :861–871, 2007. 10, 11

J.W. Philbeck and J.M. Loomis. Comparison of Two Indicators of Perceived Egocentric Distance under Full-Cue and Reduced-Cue Conditions. *Journal of Experimental Psychology : Human Perception and Performance*, 23(1) :72–85, 1997. 2, 11, 12

J.M. Plumert, J.K. Kearney, J.F. Cremer, and K. Recker. Distance Perception in Real and Virtual Environments. *ACM Transactions on Applied Perception (TAP)*, 2(3) :216–233, 2005. 17, 19, 44

V. Pulkki. Virtual Sound Source Positioning Using Vector Base Amplitude Panning. *The Journal of the Audio Engineering Society*, 45(6) :456–466, 1997. 23

L. Rayleigh. On Our Perception of Sound Direction. *Philosophical Magazine*, 13 :214–232, 1907. 6, 7

M. Rébillat, E. Corteel, and B.F.G. Katz. SMART-I^2 : "Spatial Multi-User Audio-Visual Real-Time Interactive Interface". In *Proc. of the 125th AES Convention*, number 7609, Oct. 2008. 26

A.R. Richardson and D. Waller. The Effect of Feedback Training on Distance Estimation in Virtual Environments. *Applied Cognitive Psychology*, 19(8) :1089–1108, 2005. 44

L.D. Rosenblum, C. Carello, and R.E. Pastore. Relative Effectiveness of Three Stimulus Variables for Locating a Moving Sound Source. *Perception*, 16(2) :175–186, 1987. 8

F. Rumsey. Spatial Audio and Sensory Evaluation Techniques—Context, History and Aims. In *Proc. of Spatial Audio and Sensory Evaluation Techniques Conference*, UK–Guilford, 2006. 23

B.R. Shelton and C.L. Searle. The Influence of Vision on the Absolute Identification of Sound-Source Position. *Attention, Perception & Psychophysics*, 28(6) :589–596, 1980. 1, 14

BG Shinn-Cunningham. Distance Cues for Virtual Auditory Space. In *Proc. of the IEEE Pacific-Rim Conf. on Multimedia*, pages 227–230, Sydney, Australia, 2000a. Citeseer. 4, 28

B.G. Shinn-Cunningham. Learning Reverberation : Considerations for Spatial Auditory Displays. In *Proc. of the Int. Conf. on Auditory Displays (ICAD)*, pages 126–134, 2000b. 6

JM Speigle and JM Loomis. Auditory Distance Perception by Translating Observers. In *Proc. of IEEE 1993 Symposium on Research Frontiers in Virtual Reality*, pages 92–99, San Jose, CA, USA, Oct. 25–26 1993. 3

C. Spence. Audiovisual Multisensory Integration. *Acoustical Science and Technology*, 28 (2) :61–70, 2007. 12

S. Spors and J. Ahrens. Spatial Sampling Artifacts of Wave Field Synthesis for the Reproduction of Virtual Point Sources. In *126th AES Convention*, 2009. 26

S.S. Stevens. On the Psychophysical Law. *Psychological Review*, 64(3) :153–181, 1957. 3

S.S. Stevens and M. Guirao. Loudness, Reciprocality, and Partition Scales. *The Journal of the Acoustical Society of America*, 34(9B) :1466–1471, 1962. 5

R. Teghtsoonian and M. Teghtsoonian. Range and Regression Effects in Magnitude Scaling. *Attention, Perception, & Psychophysics*, 24(4) :305–314, 1978. 3

T.V. Tran, T. Letowski, and K.S. Abouchacra. Evaluation of Acoustic Beacon Characteristics for Navigation Tasks. *Ergonomics*, 43(6) :807–827, 2000. 35

P. Willemsen and A.A. Gooch. Perceived Egocentric Distances in Real, Image-Based, and Traditional Virtual Environments. In *Proc of the IEEE Virtual Reality Conf.*, pages 275–276, 2002. 21

P. Willemsen, A.A. Gooch, W.B. Thompson, and S.H. Creem-Regehr. Effects of Stereo Viewing Conditions on Distance Perception in Virtual Environments. *Presence : Teleoperators & Virtual Environments*, 17(1) :91–101, 2008. 17, 19, 21

B. Wu, T.L. Ooi, and Z.J. He. Perceiving Distance Accurately by a Directional Process of Integrating Ground Information. *Nature*, 428 :73–77, 2004. 19

A.L. Yarbus. *Eye Movements and Vision*. Plenum press, 1967. 32

Y.Y. Yeh and L.D. Silverstein. Limits of Fusion and Depth Judgment in Stereoscopic Color Displays. *Human Factors : The Journal of the Human Factors and Ergonomics Society*, 32(1) :45–60, Feb. 1990. ISSN 0018-7208. 21

P. Zahorik. Estimating Sound Source Distance With and Without Vision. *Optometry & Vision Science*, 78(5) :270–275, 2001. 1, 15, 52

P. Zahorik. Assessing Auditory Distance Perception Using Virtual Acoustics. *The Journal of the Acoustical Society of America*, 111 :1832–1846, April 2002a. 2, 27

P. Zahorik. Auditory Display of Sound Source Distance. In *Proc. Int. Conf. on Auditory Display (ICAD)*, pages 326–332, 2002b. 24, 36

P. Zahorik, D.S. Brungart, and A.W. Bronkhorst. Auditory Distance Perception in Humans : A Summary of Past and Present Research. *Acta Acustica united with Acustica*, 91(3) : 409–420, 2005. 2, 12, 46

Z. Zhou, A.D. Cheok, X. Yang, and Y. Qiu. An Experimental Study on the Role of Software Synthesized 3D Sound in Augmented Reality Environments. *Interacting with Computers*, 16(5) :989–1016, 2004. ISSN 0953-5438. 22

C. Ziemer, J. Plumert, J. Cremer, and J. Kearney. Making Distance Judgments in Real and Virtual Environments : Does Order Make a Difference ? In *Proc. of the 3rd Symposium on Applied Perception in Graphics and Visualization*, pages 153–153. ACM, 2006. 17

Laboratoire d'Informatique des Systèmes Complexes,
E.A. 3883 (LISyC)
Centre Européen de Réalité Virtuelle (CERV)
25, rue Claude Chappe
29870 Plouzané